Beyond Inference:
Implication Models and the Future of Human-Centric AI

Michael Robbins, Washington, D.C.
www.linkedin.com/in/mwrxyz/
April 2025

Abstract

Beyond Inference: Implication Models and the Future of Human-Centric AI introduces a framework for next-generation AI systems that learn through structured implications rather than mere statistical associations. While inference-based models have yielded powerful tools, they exhibit critical shortcomings in reasoning, explainability, and alignment with human values. Implication Models address these gaps by shifting the AI learning substrate: instead of passively ingesting raw data, AI systems learn from structured representations of human experience (DOTES) that encode both what happened and what it means. The DOTES schema (Do, Observe, Tell, Explore, Show) captures experiences in a causal form amenable to machine learning, while the constructed language Mirad provides a symbolic backbone for precise, consistent concept representation.

Building on this foundation, the paper introduces AI Representatives (AI Reps): structured digital agents designed to extend and safeguard human agency, memory, and decision-making across increasingly complex digital environments. Implication Models become a cornerstone for anthropogenic AI — reframing AI as a co-evolutionary outgrowth of human culture and experience rather than an alien intelligence. This vision demands that AI be developed with people, not apart from them: built through human networks, learning ecosystems, and collaborative communities that enable authentic participation.

Implication Models offer transformative potential across education, digital networks, and governance by fostering autonomy, meaning-making, and systems that understand the world in human-relatable ways. Realizing this future will require broad interdisciplinary collaboration across the humanities, technology, and governance — weaving together education, civic society, artificial intelligence, cognitive science, linguistics, ontology engineering, game development, anthropology, philosophy, public theology, and democratic innovation to build the foundations of future digital civilization.

Author's Note (Discussion Draft – April 2025)

This paper is a discussion draft shared for review and thoughtful engagement. It is not yet finalized for publication. Please share it with others who may find it useful or wish to contribute to the conversation. Citations and excerpts are welcome, but please reference this version as a working draft.

Feedback is appreciated and can be sent to: *michael@learningpathmakers.org*

1. Introduction

Artificial intelligence has achieved remarkable successes with inference-based models that learn statistical patterns from large-scale data. Large Language Models (LLMs), for example, are trained on massive text corpora to predict the next word in a sequence, enabling fluent generation of human-like text. Similarly, image generation models like Stable Diffusion learn to map random noise to images based on patterns in billions of pictures.[1] These inference models excel at mapping inputs to likely outputs – performing what is essentially sophisticated curve-fitting on high-dimensional data. However, such systems do not truly understand the content they produce; they lack explicit representations of real-world facts, causality, or experiential meaning. As a result, they often behave as "stochastic parrots," blindly remixing correlations from their training data without any grounded comprehension or reasoning.[2] For instance, a state-of-the-art language model can generate a plausible instruction for a task yet fail to follow logical constraints or foresee the consequences of an action described in text. Likewise, an image model can render photorealistic pictures yet has no concept of the real-world dynamics or cause-effect relationships depicted. These gaps highlight a fundamental limitation of inference-based AI: association without understanding.

To reach the next stage of AI capability – human-level understanding and trustworthy autonomy – researchers are recognizing the need to incorporate causal reasoning, knowledge representation, and human values into AI systems. Judea Pearl, for example, has argued that true intelligence requires moving from "reasoning by association" to reasoning by cause and effect, enabling machines to answer not just "what is likely?" but "why?"[3] Similarly, work in neuro-symbolic AI seeks to integrate neural networks with symbolic logic to overcome the brittleness and opacity of purely statistical models.[4] These efforts reflect a broader push toward human-centered AI, which calls for AI systems that can explain their reasoning, align with human norms, and incorporate human knowledge in a meaningful way.[5]

In this paper, I introduce Implication Models as a conceptual and technical framework for learning through implications rather than inference. An Implication Model is an AI system designed to learn from experiential units that encode not only observations but the impacts and lessons of those observations. I present DOTES (an acronym for Do, Observe, Tell, Explore, Show) as a schema to structure these units of human experience into a form that an AI can ingest and reason over. Each Dotes entry (dote) captures an action (Do) and its outcome (Observe), the narrative or lesson derived (Tell), a subsequent generalization or exploration of that lesson (Explore), and a grounding in perceptual context (Show). This structured representation serves as a bridge between raw human experiences and machine-interpretable knowledge. Additionally, I employ Mirad, a constructed logical language, to encode the semantic content of Dotes entries with minimal ambiguity. By translating experiences into Mirad, I provide the AI with a consistent, symbolic substrate for reasoning, addressing the ambiguity and irregularity of natural language.

The remainder of this paper is organized as follows. **Section 2 (Related Work)** situates Implication Models in the context of prior research in AI, including inference-based learning, knowledge representation, neuro-symbolic integration, and learning from human feedback. **Section 3 (Human-Centric AI and Alignment)** explores how this approach fundamentally shifts the paradigm toward AI systems that respect human agency, digital personhood, and data dignity. **Section 4 (Methodology)** details the Implication Model framework, describing the DOTES schema and the role of Mirad in encoding knowledge, and explaining how implication-based learning operates. In **Section 5 (Technical Comparison with Inference Models)**, I provide a systematic comparison between conventional inference models and my proposed implication-based approach, highlighting differences in learning process, data requirements, knowledge generalization, explainability, and adaptiveness. **Section 6 (Human-Centered Applications of Implication Models)** discusses several domains where implication models could have high impact: education, governance, and AI alignment. I then address **Section 7 (Limitations)** of the approach, acknowledging current challenges and open questions. **Sections 8 and 9 (Near-Term Research Directions and Longer-Term Research Agenda)** outline pathways for advancing implication-based AI, from immediate implementation considerations to foundational conceptual work. **Section 10 (Anthropogenic AI and AI Representatives)** expands the vision by framing AI not as an alien intelligence but as a continuation of human evolution, emphasizing co-creation, distributed agency, and a more equitable digital economy through the development of human-aligned AI Representatives. Finally, **Section 11 (Conclusion)** summarizes my contributions and argues that moving beyond inference to implication-based AI can lay the groundwork for human-centered AI systems that learn and reason

more like humans – through understanding experiences and their consequences, not just by statistical associations.

2. Related Work

2.1 Inference-Based AI and Its Limitations

The prevailing paradigm in AI over the last decade has been dominated by *inference-based models*, which learn from large unlabeled datasets by optimizing predictive accuracy. Notable examples include deep neural networks trained for image recognition (e.g. convolutional nets on ImageNet) and large language models like BERT and GPT-3 trained on vast text corpora. These models perform **inference** in the statistical sense: given input data, they *infer* an output (a label, the next word, etc.) based on patterns learned during training.[6] Inference**,** in the context of AI deployment, refers to applying a trained model to new data to generate predictions or decisions.[7] This approach has yielded impressive capabilities. However, researchers have noted that such models often lack *robust understanding*. They are prone to spurious correlations and can fail in situations requiring reasoning not directly exemplified in the training data.[8] Bender et al. famously dubbed large language models *"stochastic parrots,"* highlighting that they can produce fluent language by recombining seen patterns without any comprehension of meaning or truth.[9] In practice, this leads to well-known issues: language models may generate factually incorrect or contradictory statements, and vision models can be easily fooled by adversarial perturbations, indicating they have not truly *grasped* the concepts but rather learned statistical shortcuts.

Another limitation of purely inference-based systems is their opacity and lack of explainability. Because their knowledge is encoded in billions of weighted connections, it is typically impossible to extract a human-understandable explanation for *why* a model gave a certain output (earning them the moniker "black boxes"). For example, an LLM might recommend a course of action in text but cannot articulate the chain of reasoning that led to that recommendation – because it has none beyond pattern matching. This is problematic for high-stakes applications that require trust and verification. As the use of AI expands to domains like law, healthcare, and governance, the demand is growing for systems that can **explain their reasoning** and ensure decisions are grounded in reliable knowledge.[10]

A third concern is that inference-trained models must often be **aligned** with human values after the fact, since training on internet-scale data introduces biases and undesirable behaviors (e.g. toxic or unsafe outputs). Techniques like *Reinforcement Learning from Human Feedback (RLHF)* have been developed to adjust model behavior by additional fine-tuning on examples of desired outputs as rated by human annotators.[11]

While RLHF and related alignment techniques (such as Anthropic's *Constitutional AI*, which guides models with explicit principles.[12]) can mitigate the worst behaviors, they do not fundamentally change how the model learns or represents knowledge; they are essentially patches on top of a system that remains a statistical learner at its core. This reactive alignment process can be brittle and may not generalize well beyond the scenarios anticipated by the fine-tuning.

2.2 Knowledge Representation and Symbolic AI

Before the rise of deep learning, AI research from the 1970s through the early 1990s was dominated by **symbolic knowledge representation**—the effort to encode facts about the world using formal structures such as logic, semantic networks, and frames. These systems relied on inference engines to manipulate symbolic expressions and derive new conclusions. A landmark project of this era was **Cyc**, initiated by Douglas Lenat in 1984, which ambitiously aimed to build a comprehensive ontology of common-sense knowledge by manually encoding millions of logical assertions.[13] The goal was to equip AI systems with a foundational real-world understanding that would enable reasoning in novel and ambiguous situations. While Cyc demonstrated that large-scale symbolic knowledge bases could support complex reasoning, it also highlighted significant challenges—including the **knowledge acquisition bottleneck** and the brittleness of symbolic systems when faced with incomplete or ambiguous real-world data.[14] These limitations eventually spurred a shift toward data-driven learning approaches in AI.

The **symbol grounding problem** is a fundamental issue that symbolic AI grappled with: how to connect abstract symbols and logical expressions to real-world meaning.[15] Stevan Harnad famously argued that purely symbolic AI systems risk being a "mere manipulation of symbols" without genuine understanding — analogous to Searle's **Chinese Room** thought experiment, where a person follows syntactic rules to manipulate Chinese characters without actually knowing the language.[16] In other words, for symbols to be meaningful, they must ultimately relate to perception or experience. This realization has driven increasing interest in **multi-modal grounding**—integrating images, audio, or sensor data with symbolic knowledge so that AI systems can link symbols to real referents. Some modern approaches combine vision and language models to create **joint embeddings** that map textual descriptions and visual elements into shared spaces, partly addressing grounding through multi-modal training. However, true grounding likely requires AI systems to learn through **interaction with the world or with detailed simulations of experience**, rather than static association alone. For instance, **Joint Embedding Predictive Architectures (JEPA)**, proposed by Yann LeCun, aim to move beyond supervised learning by enabling AI systems to predict and reason about future sensory inputs based on world models learned through interaction.[17]

2.3 Learning from Experience and Cases

Humans learn not only from passive observation or instruction, but critically from *direct experience*: I perform actions, observe the outcomes, reflect, and adjust my understanding. In cognitive science and education theory, **experiential learning** is known to produce deep understanding. Kolb's experiential learning model (1984) describes a cycle of concrete experience, reflective observation, abstract conceptualization, and active experimentation.[18] Essentially, learning by doing and then thinking about what was done yields new generalizations that can be tested in practice. This has clear parallels to how one might design an AI that learns like a human – by having it *encounter* scenarios through human-provided narratives and derive lessons.

Earlier AI paradigms like **Case-Based Reasoning (CBR)** also emphasized learning from concrete examples or "cases." A case-based reasoning system stores a library of past cases (problems and their solutions or outcomes) and, when faced with a new problem, retrieves similar cases to suggest a solution by analogy.[19] CBR thereby uses *experiences* (in the form of cases) as the primary knowledge resource, rather than abstract rules. Notably, CBR systems often include an explanation component, since a retrieved case serves as an explicit precedent: the system can say "I propose solution X because in a similar past case Y, that solution was successful."[20] This is an early example of an AI reasoning via implications of previous events (if situation Y implied solution X was good, perhaps current situation Y' will imply X' is good). However, traditional CBR systems required well-structured cases and did not automatically learn the underlying principles; they were also limited by the contents of their case libraries and struggled if no sufficiently similar case was available.

2.4 Neuro-Symbolic Integration

Recent years have seen a resurgence of interest in combining neural and symbolic methods – often termed **neuro-symbolic AI** – to get the best of both worlds.[21] Neural networks are excellent at pattern recognition and handling noise in high-dimensional data, whereas symbolic approaches excel at representing explicit knowledge and reasoning with logical precision. By integrating the two, researchers aim to create systems that can learn from raw data *and* manipulate abstract concepts. For example, neuro-symbolic systems have been developed for reasoning over **knowledge graphs**, where a neural model might learn to embed entities and relations, but a symbolic reasoner can still perform logical queries over the graph.[22] One survey highlights that neuro-symbolic AI can achieve improved generalization from fewer examples, by leveraging prior knowledge and structure.[23] In essence, a neuro-symbolic system can use neural components to interpret inputs (e.g. image recognition, language parsing) and then use symbolic components to reason about these inputs in a human-like way (e.g. executing a logic rule or a relational query).

This line of work supports the intuition that moving beyond pure inference requires an *architectural change* in AI systems. Rather than a single black-box model, a combination of subsystems – some learned, some engineered – working together can yield an AI that is both proficient at pattern matching and at reasoning. My proposed Implication Model framework can be seen as a neuro-symbolic approach: it envisions using neural techniques to process raw sensory data (like images or free text) and using symbolic representations (DOTES entries encoded in a formal language) for higher-level reasoning and knowledge organization. The balance between learned associations and explicit reasoning is a key design question for any such system.

3. Human-Centric AI and Alignment

My work is motivated by the agenda of **human-centric AI** and AI alignment. Unlike merely *human-centered* approaches that place humans in the middle, human-centric AI puts people actively in the loop and in charge, recognizing their agency and autonomy. This paradigm acknowledges that technology should serve human flourishing, not merely accommodate human needs within systems primarily designed for other purposes.

Human-centric AI advocates for systems that fundamentally respect digital personhood and data dignity. This includes transparent reasoning processes, human-directed controllability, and fully participatory design where humans aren't merely consulted but have decisive authority in development processes. Implication Models embody these principles by using *human experiences* as their primary training data—placing human knowledge, values, and wisdom at the core of the AI's understanding.

Rather than extracting data from the internet without meaningful consent (the status quo for many models), implication-based AI learns from stories, experiments, and demonstrations willingly contributed by users. This ensures the knowledge base is grounded in human contexts with clear provenance and attribution for each knowledge element. By recognizing data as an extension of personhood deserving dignity and respect, this approach creates systems inherently aligned with human sovereignty.

This framework democratizes AI development: as diverse individuals contribute their unique experiences, the resulting AI draws from a collective pool of human wisdom representing multiple perspectives, cultures, and knowledge traditions. This stands in stark contrast to current models biased toward predominantly Western, internet-popular content. When AI systems recognize and respect the dignity of human data providers, they become genuine partners in advancing human potential rather than tools that merely extract value from human information.

In terms of alignment, building an AI's understanding through human-curated experiences offers a novel path to imbue human values while respecting individual

autonomy and agency. This human-centric approach positions AI as augmenting rather than replacing human capacity, creating a symbiotic relationship that upholds human flourishing as the ultimate measure of technological success.

In terms of alignment, building an AI's understanding through human-curated experiences offers a novel path to imbue human values. For example, experiences that encode moral lessons (a story of why lying caused harm or how helping someone had positive outcomes) would directly teach the AI representations of those principles in context. This is a different paradigm from methods like RLHF where values are imposed on a pre-trained model via reward signals.[24] Instead, values and norms could be *intrinsically learned* by the Implication Model as it generalizes from the experiences provided. Moreover, because the knowledge is stored in an interpretable format, alignment researchers or ethicists could inspect and audit the AI's "mind" – essentially the corpus of Dotes – to see what it has learned and whether that aligns with desired ethics. Anthropic's **Constitutional AI** approach, which uses a fixed set of written principles to guide model outputs, demonstrates that even simple, explicit rules can significantly shape model behavior.[25] Implication Models extend this idea by allowing AI to learn a *rich set of nuanced principles and heuristics* through example and analogy, rather than being limited to a static list of rules. This could address subtle behaviors and context-dependent judgments that are hard to encode in universal rules but can be illustrated via scenarios.

In summary, the landscape of AI research provides several building blocks and motivations for my work: the shortcomings of inference models highlight what needs improvement; the legacy of symbolic AI and modern neuro-symbolic successes suggest the power of structured knowledge; and human-centered AI principles urge us to design systems that learn in concordance with human ways of knowing. Implication Models aim to synthesize these threads into a coherent approach, which I detail next.

4. Methodology: DOTES and Implication Models

In this section, I present the methodology for building *Implication Models* grounded in human experiences. First, I define the concept of an Implication Model more formally and contrast its learning objective with that of traditional inference models. I then introduce **DOTES (Do, Observe, Tell, Explore, Show)**, a schema for encoding experiences in a structured, multi-faceted way that captures both the factual and contextual aspects of learning from those experiences. I describe each component of DOTES and how it contributes to an AI's ability to reason about implications. I introduce a taxonomy that categorizes experiences by core domains of human development. Next, I discuss how these Dotes entries are represented symbolically, with a focus on the use of the constructed language **Mirad** to encode meanings precisely. Finally, I outline how an

AI system would be trained and operate using Dotes data – that is, how implication-based learning is achieved in practice.

4.1 From Inference to Implication: A New Learning Paradigm

An Implication Model is, at its core, an AI model that learns to anticipate and reason about the implications of actions and events, rather than merely predicting correlations in data. Unlike traditional inference models, which rely on statistical pattern recognition to forecast likely outcomes, Implication Models integrate structured representations of cause and consequence. For example, consider a scenario where a person kicks a ball toward a window. A standard inference model might predict that the window breaks because it has observed similar co-occurrences in training data. An Implication Model, by contrast, reasons that kicking the ball → ball hits the window → glass shatters → person gets in trouble. It understands not only what is likely to happen, but why — and what it might mean for future behavior. This distinction places emphasis on consequence chains, not just outcomes, and supports learning that resembles narrative reasoning or experiential learning rather than pure statistical extrapolation.

Formally, we can think of implication learning as learning a function (or a set of functions) F such that from an input situation or event X, the model produces not only an output Y, but also a chain or network of inferred outcomes/implications $\{I_1, I_2, ..., I_n\}$ that stem from X. This can include immediate effects, longer-term consequences, and generalized lessons. For example, if X = "a user touches a hot stove," an implication-aware model would infer Y = "the user gets burned" and might further infer I_1 = "the user feels pain," I_2 = "the user learns a lesson to avoid touching hot stoves," and possibly generalize I_3 = "hot objects can cause injury." By contrast, an inference model might only learn a statistical association between the words "touches a hot stove" and "burn," without any chain of reasoning or abstraction. Implication Models thus aim to answer questions such as: "If X happens (or is done), what are the consequences, and what does it imply for future choices?"

To enable such learning, I propose a knowledge representation and training data format specifically tailored to capture experiences and implications. That is the role of **DOTES**.

4.2 The DOTES Schema: Encoding Human Experience

DOTES is a five-part schema for encoding experience in a structured, causal, and interpretable format. The components — **Do, Observe, Tell, Explore, Show** — represent distinct phases of human reflection and learning, designed to preserve both the concrete reality of what occurred and the subjective process of making sense of it. A Dotes entry may represent a real or imagined event, but it is always expressed from the first-person perspective of the individual who experienced it. Each component contributes a layer of meaning:

Do – *Summarize your actions or experience*. This component captures the initiating action or event. It is expressed objectively from the participant's point of view: what was done, attempted, or encountered. For example: "I hit a beehive with a stick." The Do component anchors the experience in a specific behavior or moment and serves as the basis for causal interpretation.

Observe – *What happened as a result? What did you notice?* This component records the outcomes that followed the action — both external consequences and internal states. It may include physical effects, emotional reactions, and environmental changes. For instance: "Bees flew out of the hive. I felt a sharp pain in my arm and saw redness where I was stung."

Tell – *What did you take away from the experience?* This captures meaning-making and narration. It may include internal reflections, explicit lessons, or feedback received from others. The Tell provides interpretive context: "My teacher said the bee was defending the hive. She explained that bees die after stinging. I realized the bee wasn't attacking — it was protecting."

Explore – *What will you do next time? What do you still wonder about?* The Explore component encodes forward-facing thinking. It may reflect a behavioral intention, a hypothesis, or a question that arises from the experience: "Next time, I'll leave bees alone. I want to learn more about why they sting and how they live."

Show – *What visual or sensory artifact grounds this memory?* The final element connects the experience to a perceptual referent — a photo, video, sketch, or sound — that provides symbolic grounding. For example, this Dotes entry might include an image of a honeybee and a beehive to reinforce the subject visually and reduce ambiguity in interpretation. By linking abstract components of the experience to concrete sensory data, Show addresses the symbol grounding problem in AI — the challenge of connecting internal representations to the external world — and ensures that experiences are encoded not only semantically but perceptually.[26]

Together, these five components comprise a Dotes entry: a coherent, multi-dimensional record of experience. The structure allows for systematic parsing, indexing, and generalization while preserving the nuance and meaning embedded in human memory. Each component contributes to forming a high-resolution representation that is both reflective for humans and tractable for machines. Later sections will illustrate how such structured entries support causal modeling, implication-based learning, and explainable AI.

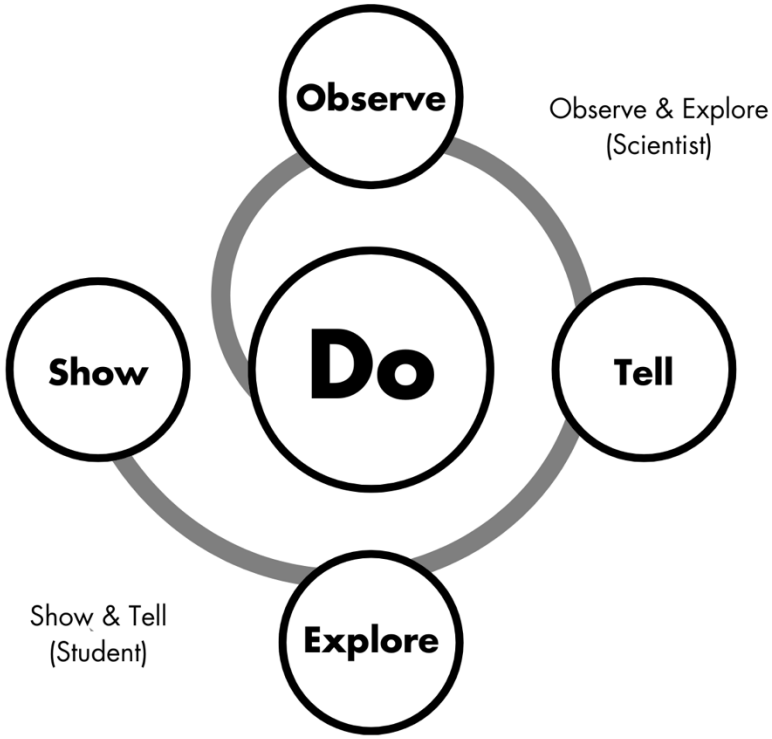

Figure 1. *The DOTES framework connects early narrative forms of learning (e.g., "Show and Tell") with later scientific practices (e.g., "Observe and Explore"), highlighting a developmental continuum from intuitive storytelling to formal inquiry. This spiral structure reflects how learners of all ages engage in reflective cycles of action, observation, explanation, and iteration.*

This framework parallels and extends **Kolb's Experiential Learning Cycle** which consists of four iterative stages: concrete experience, reflective observation, abstract conceptualization, and active experimentation.[27] The DOTES schema aligns with this model but adds critical structure and multimodal expressivity. *Do* corresponds to the concrete experience; *Observe* reflects on outcomes; *Tell* encodes abstracted meaning or generalization; and *Explore* initiates future application or inquiry. *Show*, the fifth component, introduces a sensory or contextual anchor not formally captured in Kolb's model, enabling symbolic grounding and visual disambiguation critical for machine interpretation. In this way, DOTES serves as both a pedagogical extension and a computational formalization of experiential learning — one capable of bridging human reflection and AI representation.

Example: "Don't Provoke Bees" – A DOTES Illustration

To make the concept concrete, consider a modified version the earlier example formulated as a first-person Dotes entry:

> **D (Do):** I disturbed bees. A bee stung me.
>
> **O (Observe):** My arm had a sharp shooting pain. I looked at it and saw something was stuck in my arm and my arm was red.
>
> **T (Tell):** There were bees in a bush. I hit the bush with a stick. Ouch! I had a sharp shooting pain in my arm that kept hurting. I ran over and told the teacher. The teacher said it was a bee sting. The bee stung me and left its stinger in my arm. It was protecting the other bees. My teacher told me the bee died after that.
>
> **E (Explore):** Whenever I see bees, I will be careful. Give them space. Don't mess with bees.
>
> **S (Show):** *(Image of a honeybee and a beehive).*

In this package, an AI doesn't just see the components as separate pieces of text; it sees the full narrative with an explicit causal link and an articulated implication ("Don't mess with bees"). If this is one entry among many in the AI's training corpus, the AI can begin to build a **cause-effect knowledge graph**. For instance, it might link the concept of *"disturbing bees"* to *"getting stung"* with a relation *causes* and also link *"getting stung"* to *"feeling pain"* (from O), and link *"don't disturb bees"* as a recommended rule arising from that situation. Over multiple such entries, patterns emerge: e.g., several experiences might involve *pain* as an outcome of certain actions, teaching the AI a general concept that pain indicates a bad outcome to be avoided.

This approach bears some resemblance to how children learn: through **stories and fables** that explicitly come with morals, or through personal experience followed by guidance from parents and others. By encoding Dotes entries from many people, we accumulate a wide-ranging database of "small stories" each with a moral or implication. In essence, the Implication Model's knowledge base is a collection of *experiential narratives* rather than a collection of isolated data points or static facts.

4.3 Taxonomy of DOTES: Connecting Experience to Human Purpose

While the DOTES schema structures how individual experiences are recorded, we introduce a complementary **taxonomy** to categorize Dotes by their thematic focus. This taxonomy anchors learning experiences within broader domains of personal and

communal development, facilitating reflection, retrieval, and implication modeling across common patterns of growth.

The taxonomy is organized into seven developmental categories, presented in a proposed sequence for common use in applied settings:

1. **Character** – choices and decisions
2. **Excellence** – proud achievements
3. **Service** – helping others and the community
4. **Relationships** – building connections.
5. **Adventure** – exploring new things.
6. **Making** – creativity and creation.
7. **Wellness** – embracing health and balance.

Each dote may be tagged with one or more categories depending on its core intention and outcome. This taxonomy enhances the semantic richness of DOTES-structured memories by providing a **higher-level scaffold** for clustering experiences and reasoning about patterns of growth over time.

By aligning AI memory structures to these human-centric developmental domains, the system can not only predict outcomes but also discern the broader *purpose* and *growth vector* of actions — a critical step toward human-aligned, experience-grounded AI.

4.4 Symbolic Representation with Mirad for Machine Comprehension

One of the challenges in implementing the above ideas is ensuring that the AI can robustly parse and reason over the content of Dotes entries. If we were to use raw natural language (e.g. English) for the D, O, T, E descriptions, we risk reintroducing ambiguity and complexity. Natural language, while rich, is full of irregularities and context-dependencies (e.g., the word "bat" could mean a flying mammal, or a club used in sports). If the AI tries to learn directly from a large number of English narratives, it might struggle to differentiate nuances or might incorrectly generalize due to linguistic ambiguities. Moreover, the grammar and phrasing variations in natural language could add noise to the learning process.

To address this, I propose using a constructed language — specifically, Mirad — as an intermediary symbolic representation for the content of Dotes entries. Mirad (formerly known as Unilingua) was originally created by Noubar Agapoff of Paris in 1966. It was later discovered, translated from French, and substantially modernized by Jamie Shoemaker, a former linguist at the NSA, who renamed it from Unilingua to the more internally consistent term Mirad (Worldspeech). Shoemaker made significant

improvements to the language, including replacing the mixed Latin-Cyrillic alphabet with an all-Roman system, discarding the noun case system in favor of prepositions, and expanding the systematic vocabulary to over 200,000 word/expression pairings using fewer than 500 root words. This modernized version has been documented in a publicly available Wikibook.[28]

Mirad is designed to be **regular, logical, and unambiguous**. Its vocabulary and grammar are constructed in a way that each element carries clear semantic or grammatical meaning (for instance, each vowel and consonant has a systematic role). Words in Mirad are built from a structured ontology, making it a *taxonomic language* – words with related meanings share common roots or patterns, and there are minimal idiomatic exceptions. The author of Mirad specifically intended it to be a language optimized for logical communication, akin to how mathematical notation is a universal, unambiguous language for quantitative concepts.[29]

For example, in Mirad, the sentence "Don't mess with bees" can be rendered precisely and without ambiguity. The translation is: "Von loboxu appelati" (where *loboxu* means "to disturb" and *appelati* means "bees"). Because Mirad has a consistent and compositional structure, an AI can parse this sentence and understand its components far more easily than an English equivalent—which might include idiomatic phrasing, cultural nuance, or crass colloquial slang that springs to mind while describing a painful encounter.

By translating the **D, O, T, E textual components** of each Dotes entry into Mirad (or by initially recording them in Mirad), we give the AI a uniform, logically structured dataset. This could be seen as analogous to how computer programs often use an intermediate representation or a canonical form of data for internal processing. Mirad serves as a *knowledge representation language* for our AI. It is human-readable (for those who learn it) but more importantly, it is *machine-friendly*. Key advantages of using Mirad include:

- **Elimination of Ambiguity:** As noted, Mirad strives to have one word for one concept, and maintain distinctions clearly. Words are "ontologically unambiguous". This means the AI is less likely to confuse terms or misconstrue a sentence's meaning. For instance, English might use the word "sting" both as a noun (the sting of a bee) and a verb (the bee stings). Mirad might use distinct forms that make the grammatical role clear.
- **Consistency in Grammar:** Mirad has regular grammar rules without exceptions. This regularity means an AI can easily parse sentences – it doesn't need complex machine learning just to understand the syntax, unlike English where I often use sequence models to parse because of

irregular structures. A context-free grammar for Mirad can be coded by experts and reliably used by the AI.

- **Compositional Semantics:** Mirad words are constructed such that similar concepts share roots (for example, if *iva* means happy and *uva* means sad, that systematic vowel change indicates opposition). This built-in structure could help the AI to generalize; learning one word might immediately help it recognize related words. It also can reduce the memory burden – the AI might infer meaning of a new Mirad word from its composition, rather than treating every word as an unrelated token.
- **Mapping to Symbolic Structures:** Because Mirad is designed like a logical or mathematical language, sentences in Mirad can often be mapped to formal logic or semantic graphs relatively straightforwardly. This could facilitate an automated conversion of Mirad-encoded knowledge into a knowledge graph format or into predicate logic for reasoning. In effect, Mirad can act as a high-level interface for encoding knowledge that is then stored in a semantic network within the AI.

It's important to note that **Mirad is an auxiliary tool**; the concept of Implication Models does not strictly depend on Mirad specifically, but on having a structured representation. I could have alternatively chosen other controlled natural languages or even directly a logical formalism. I chose Mirad due to its balance of readability and structure – it's a full language (so it can express nuances of experiences in a relatively compact human-like form) while being much more regular than natural languages. Prior work has shown that using *interlingua* or constructed languages can help in machine translation and comprehension tasks because it reduces complexity for the model. Here, Mirad plays the role of an interlingua between the human contributor of a Dotes entry and the AI's internal reasoning system.

Table 1: DOTES English-Mirad Side-by-Side Translation

DOTES Component	English	Mirad Translation
Do	I disturb a bee. The bee stings. I hurt.	At loboxe appelat. Ha appelat vuloxe. At byoke.
Observe	Pain. The bee stinger is stuck in my arm. The bee flies away.	Byok. Ha appelat vulob se kyoxwa yeb ata tub. Ha appelat papie.

DOTES Component	English	Mirad Translation
Tell	The tree is big. The bee is near. I didn't know. I touched it with my arm. I hurt. I ran and told the teacher. The teacher said the bee didn't like it. The bee protected other bees. The bee stung and left its stinger in my arm. The bee died.	Ha fab se aga. Ha appelat se yuba. At voy ta. At byuxa it bay ata tub. At byokia. At igtyopa ay da ha tuxut. Ha tuxut da van ha appelet voy iyfa his. Ha appelat ovmasba hyua appelati. Ha appelat vuloxa ay ba ita vulob yeb ata tub ay ipa. Ha appelat toja.
Explore	Whenever I see a bee then I must be careful. Give space. Don't disturb bees.	Hyej at teate appelat at yefe bikier. Buu nig. Von loboxu appelati.
Show	[Image of a honeybee and a beehive]	[Tagged as image_bee_hive_001]

Note: This translation is for illustration of the Mirad encoding concept. The Mirad vocabulary demonstrates how a constructed logical language can represent experiences with minimal ambiguity.

Key Mirad Vocabulary Elements in This Example:

- **appelat**: bee
- **loboxe/loboxu**: disturb (verb form variations)
- **vuloxe**: sting (action)
- **byoke/byokia**: hurt/pain (variations)
- **tuxut**: teacher
- **von**: negation marker ("don't")

The Mirad text would be stored along with perhaps an English annotation for human developers to cross-check. The AI, during training, would process the Mirad version as primary input. It might learn embeddings for Mirad tokens that capture their semantic relationships (e.g., embedding for “appelati” (bee) close to “insect”, etc., if it has the ontology encoded). When reasoning or answering questions, the AI could map back from Mirad to natural language for human output or internally think in Mirad-like structures. In essence, Mirad provides a **“language of thought”** for the AI that is far more constrained and algebraic than natural language, which is beneficial for reliable reasoning.

The use of Mirad connects to the notion of **symbolic alignment**: aligning the AI's internal representations with human-meaningful symbols. Since Mirad is human-designed but also machine-friendly, it serves as a middle ground. By training the AI on Mirad-encoded knowledge, we avoid the AI developing *entirely opaque internal representations*; instead, its internal structures are themselves in a language we (at least theoretically) understand. This contributes to explainability: when the AI provides an answer or rationale, it could be prompted to output the supporting Dotes entries (which are in Mirad but could be translated to English and other languages) or to output a logical chain in Mirad that can be interpreted.

4.5 Training an Implication Model

How would we train an AI on DOTES data in practice? This is an important implementation question. While a full engineering solution is beyond the scope of this paper, I outline a possible approach:

1. **Data Collection:** Assemble a corpus of Dotes entries. This could involve a platform where users contribute experiences in a structured format (first in natural language, then translated). Ensuring quality and diversity is important – each entry should ideally be vetted for correctness (the stated implications should logically follow from the described events).
2. **Encoding:** Translate or encode all textual components of the Dotes entries into Mirad. Validate that the Mirad accurately captures the intended meaning. For the Show component, ensure images are labeled or described so the AI can connect them (e.g., perhaps by providing captions in Mirad as well, like "This is a bee").
3. **Model Architecture:** Utilize a multi-modal neural network architecture that can handle:
    - Textual input in Mirad (e.g., a transformer or recurrent network for sequences).
    - Visual input for images (e.g., a CNN or vision transformer for the Show component).
    - Possibly a graph neural network or memory network to store and relate multiple Dotes entries. One could imagine an architecture where each Dotes entry is processed, and key embeddings (for the scenario, the lesson, etc.) are stored in a memory. The model should be able to attend to relevant past experiences when answering a question or making a decision.
4. **Training Objective:** Instead of the usual next-word prediction, objectives for implication learning could include:

- **Consequence Prediction:** Given D (and maybe S), predict O (this trains cause-effect understanding).
- **Lesson Identification:** Given D and O, output the appropriate T (the model learns to articulate the implication).
- **Generalization:** Given D (and O), predict E (what general rule or future behavior should apply).
- **Consistency Check:** Ensure that applying the E rule to D would prevent O (in cases where O was negative). For instance, if E says "avoid doing D", then indeed D caused a bad O.
- **Question Answering:** On held-out scenarios, ask the model questions like "What should one do in situation X?" or "Why did Y happen after X?" and train it to answer using the implications learned. These objectives encourage the model to internalize the relationships within each Dotes entry and across them. It's not just learning to mimic text, but to predict *structured outcomes* and *abstract lessons*. This is closer to how one might train a model to do theorem proving or planning, rather than free-form text generation.

5. **Reasoning Mechanism:** During inference (when the model is deployed), it would use its learned knowledge to reason about new inputs. For example, if asked a question in English, the system could translate the question (or key parts) into Mirad, query its memory of Dotes for related experiences, and then compose an answer. An advantage of having explicit Dotes entries is that the model can do a kind of case-based reasoning: find similar experiences to the query and use their lessons to form an answer. Because those lessons are explicit (T/E) and understandable, the model can even quote or refer to them in its answer, providing an explanation.
6. **Feedback and Update:** As the model interacts with users or an environment, new Dotes entries can be continuously added. For instance, if the model encounters a novel situation it can't handle, a human can provide a new example as Dotes. The model's knowledge base then grows. Since knowledge in implication models is modular (each experience is a module), updating the model doesn't necessarily require retraining from scratch on a huge corpus (unlike current LLMs which need to be retrained or fine-tuned on new data). Instead, one can insert new Dotes and perhaps fine-tune the model's representations slightly or allow it to attend to the new entries. This could make learning *dynamic and lifelong*, an essential feature of any system that aims to be *human-like in learning*.

By combining the structured Dotes data with neural learning and symbolic reasoning, the AI develops what we can call an **experiential knowledge base**. It is *neuro-symbolic*: neurons (or embeddings) capture the patterns, but symbols (Mirad and the Dotes structure) ensure those patterns align to meaningful concepts and relations. In a sense, we are asking the model to **construct a world model** – a mental model of how the world works – from the ground up using human experiences as building blocks. This is reminiscent of the concept of *common-sense knowledge* in AI. Instead of hoping the model *emerges* common sense by reading internet text, we explicitly feed it digestible pieces of common sense (experiences with their common-sense lesson). The approach also echoes Judea Pearl's advocacy for causal models: we are giving the AI data points that include interventions (Do) and outcomes (Observe), essentially guiding it to learn a causal graph of events.

I stress that my proposed methodology is *hybrid*: it does not throw away the progress of statistical learning but rather directs it. The AI still uses pattern recognition to handle the perceptual aspects (e.g., identifying the bee in the image, parsing the Mirad phrases), but it augments that with a **cause-effect inference engine** for the higher-level implications. The result is an AI that can answer not only "What likely comes next?" but also "*Why did this happen?*" and "*What should be done?*", drawing upon its library of learned experiences.

4.6 DOTES and the Conveyance of Tacit Knowledge Through Stories

A core strength of the DOTES schema lies in its capacity to convey **tacit knowledge**—the kind of understanding that resists codification but shapes how people act, decide, and relate. Tacit knowledge isn't just what we know, it's what we *carry*—what we've internalized through doing, reflecting, and learning over time. It doesn't come from manuals or rules. It comes from experience. It lives in stories.

DOTES was designed with this in mind. Every entry is a microcosm of lived meaning: an action, an outcome, a reflection, a lesson, and a sensory anchor. Together, they form a kind of structured anecdote—a **micro-parable**. This isn't just data for machines. It's a format that lets people preserve and share their own hard-won insights, in a way that retains their full narrative integrity. It's structured but not sterilized.

Across history and culture, **stories have always carried the weight of what mattered**. Parables, fables, myths, and cautionary tales weren't just entertainment—they were memory systems. They taught not only what to do, but what to value. From the Bhagavad Gita to the Book of Proverbs, from Yoruba folktales to Zen koans, human beings have used narrative to convey not just knowledge, but *wisdom*.

DOTES offers a modern scaffold for this ancient function. By encoding experiences in this format, we aren't just creating better training data for AI. We're creating a ledger of meaning that can be passed from person to person, generation to generation, machine to machine—with the full dimensionality of human understanding intact.

And not just understanding. Wisdom, by its nature, resists simplification. It often appears as intuition—a kind of knowing that can only be applied by someone who understands their own context. Two people can hear the same story and draw different implications, depending on where they are in their journey. DOTES preserves that possibility. It doesn't prescribe a universal lesson; it presents an experience, complete with consequences and reflections, and leaves space for human interpretation. The model isn't meant to replace judgment—it's meant to help people encode the patterns, so they can apply them with care.

5. Implication Models vs Inference Models

To better understand the impact of the Implication Model approach, I compare it against existing inference-based models along several key dimensions:

Table 2: Comparison of Inference Models and Implication Models

Dimension	Inference Models	Implication Models
Training Data and Knowledge Source	Trained on massive unstructured datasets, often scraped without consent. Knowledge is inferred and untraceable.	Trained on structured, human-curated Dotes entries. Knowledge is explicit, traceable, and contributed with consent.
Learning Mechanism	Learn patterns probabilistically by minimizing prediction error; no distinction between causation and correlation.	Combine symbolic reasoning with adaptive inference; learn cause-effect chains and generalizable rules.
Knowledge Organization	Knowledge is embedded in model weights; difficult to update or inspect.	Knowledge is modular and stored externally (e.g., as Dotes or graphs); can be updated without full retraining.

Dimension	Inference Models	Implication Models
Multi-modal Grounding	Typically, modality-specific or aligned statistically; lacks grounding in causality or sensory experience.	Inherently multi-modal with symbolic grounding via Show components; grounded in perception and meaning.
Generalization and Adaptability	Generalize by interpolation; struggle with novel situations and require retraining to incorporate new knowledge.	Capable of abstraction and analogical reasoning; learn lessons that generalize to novel or related cases.
Explainability	Opaque decision-making; explanations are generated post hoc and not traceable to internal reasoning.	Reasoning is traceable and transparent; models can cite source experiences and rule chains for outputs.
Goal Orientation (Proactivity vs. Reactivity)	Reactive to input prompts; no foresight or awareness of implications.	Designed to anticipate outcomes and make decisions with foresight; embed proactive, ethical behavior.

By comparing these, we see that implication models address many weaknesses of current AI. They are deliberately *designed* for alignment, interpretability, and **common-sense reasoning**, whereas those are afterthoughts or ongoing challenges in inference models.

Of course, it must be acknowledged that inference models have one strength: sheer performance in pattern matching due to scale. Implication models, especially in their early stages, might not match the raw fluency or perceptual accuracy of a model that has digested terabytes of data. However, one can envision a hybrid: using a pre-trained inference model as a component within an implication model system. For example, an LLM could be used to help parse user input into a DOTES-style query or to generate a draft Tell given a Do and O, which the implication system then checks and refines. Ultimately, as implication models gather more data (since human experience is endless and diverse), they could approach the breadth of knowledge of current models, but with far superior structure. In the long term, the trade-off favors implication models for any application where correctness, reasoning, and alignment matter more than surface-level fluency.

6. Human-Centered Applications of Implication Models

Beyond their technical underpinnings, Implication Models have broad significance for how AI systems interface with human society. This section explores how learning from structured experiences extends into key human domains. By integrating *experiential* knowledge into AI, we can influence education, governance, personal identity development, and workforce evolution. These areas illustrate that the value of Implication Models lies not just in algorithmic performance, but in fostering AI that grows with and alongside people.

6.1 Implication Models for Education

Traditional education often separates formal learning from the realities of everyday life, focusing on abstract concepts detached from the lived experiences that shape personal growth. Implication models offer a new paradigm: AI systems that can learn from and reason about human experiences captured through structured reflection. By aggregating and reasoning over Dotes — real stories of action, observation, insight, adaptation, and sensory grounding — an AI tutor can support learners not only in mastering academic content but also in navigating complex life domains.

6.2 Implication Models for Governance

In governance, policymakers equipped with implication-model decision support might review aggregated first-person case studies (rather than just statistics) to understand and anticipate the human impact of policies. By reasoning from historical and grassroots experiences, such systems could enhance transparency and trust in decision-making, grounding policies in lived consequences rather than abstract models.

6.3 Implication Models for AI Alignment and Safety

Another crucial domain is **AI alignment**. Training AI on human-curated experiences instills a form of moral and common-sense grounding that pure data-driven training lacks. Whereas conventional AI might correlate inputs and outputs without context, an implication model learns the *meaning* behind actions – effectively absorbing a "moral compass" from cumulative human lessons. This experiential grounding could help autonomous systems make decisions that align better with human values and societal norms. In essence, across these domains the common thread is that AI systems become partners in *experiential learning* – they learn *with* us by understanding the consequences that we care about, rather than only optimizing isolated metrics. Implication Models thus act as bridges between technical systems and human values, making AI's knowledge more context-rich and its actions more accountable to real-world outcomes.

6.4 Narrative Identity, Digital Identity, and the Role of DOTES

Human identity itself can be viewed as an ongoing narrative constructed from life experiences. Psychological research postulates that individuals form their identity by integrating experiences into an internalized, evolving life story that provides a sense of unity and purpose.[31] Notably, adolescence is the formative period when one's "narrative identity" coalesces; it aligns with Erikson's observation that the central task of youth is to answer *"Who am I?"* and achieve a coherent sense of self[32] A consistent personal narrative helps link one's past, present, and future, yielding continuity across the many roles and contexts a person inhabits.

The DOTES framework can serve as a scaffold for this narrative identity in the digital age. By encouraging individuals to **Do, Observe, Tell, Explore, and Show** their experiences, DOTES creates a structured reflective diary of life events and lessons. Over time, such a corpus becomes a form of **digital identity** or "digital personhood" – a curated narrative of one's growth, values, and knowledge encoded in data. Rather than disparate social media posts or static profiles, a DOTES-based personal archive would emphasize coherence and meaning. This supports healthier identity development: the individual (especially an adolescent) can revisit and make sense of their experiences across domains (school, family, online) with the AI's help, reinforcing a stable yet evolving self-story. The AI, in turn, uses this narrative to personalize its interactions, treating the user as a whole person with history and goals, not just a set of queries. In short, implication-oriented systems promote *narrative identity formation* by linking episodes into lessons – helping people reflect on who they are across time and digital spaces and empowering them with a richer understanding of their own journey.

6.5 DOTES for Employment and Workforce Development

The modern workforce is characterized by rapid skill turnover and the need for **lifelong learning**. Continuous upskilling and adaptation have become the "new norm" if individuals and organizations want to stay ahead. However, traditional credentialing – degrees, certificates, and static resumes – often fails to capture a person's evolving capabilities. A diploma provides a snapshot of knowledge at one point in time, but it says little about the practical wisdom gained on the job or how someone's skills have grown through experience. Research indicates that accumulated work experience contributes roughly 40–60% of an individual's human capital value, emphasizing how much of our professional ability comes from learning-by-doing over years. [33] There is a clear need for more **portable, narrative-based representations** of skills and growth that workers can carry throughout their careers.

The DOTES approach offers an alternative: a living résumé of experiences and lessons that evolves with the person. By recording concrete scenarios of challenges faced, actions taken, outcomes observed, and insights gained, professionals build a rich evidence-based narrative of their competencies. Such an **experiential portfolio** could be leveraged in hiring and career development. Instead of relying solely on titles or test scores, employers could review a candidate's relevant DOTES entries to understand how they handled real situations – akin to a personal case study archive. This narrative format highlights transferable skills like problem-solving, adaptability, and teamwork in context, which static credentials may overlook. It also empowers individuals to reflect on their growth and proactively identify skill gaps to explore next. In workforce development programs, DOTES can support mentoring and training by making tacit knowledge explicit: for example, a trainee's DOTES log allows coaches to give targeted feedback on each experience.

In sum, implication models enable a shift from credential-centric to **experience-centric** workforce development. Careers become journeys of continual learning documented in narrative form, and AI systems using these narratives can more intelligently match people with opportunities, recommend learning pathways, and recognize achievements that happen beyond the classroom or certification exam. This dynamic, story-based approach to skills not only complements traditional qualifications but could eventually redefine how we assess and **credential human capability** in the age of constant change.

7. Limitations

While the Implication Model paradigm holds significant promise, it also faces several limitations and challenges. It is important to address these candidly, both to avoid overhyping the approach and to direct future research to overcome these obstacles.

7.1 Data Collection and Scalability

Building a large, high-quality corpus of Dotes entries is a non-trivial task. Unlike scraping the web for text (which is automatic but indiscriminate), compiling structured experiences requires human input and curation. This could prove to be a bottleneck. We might initially rely on experts (e.g., educators for educational content, domain experts for medical cases) to contribute Dotes, but to scale to millions of entries reflecting the broad spectrum of human knowledge, we need widespread participation. Motivating contributions (perhaps through incentives or the intrinsic appeal of contributing to a collective AI mind) will be crucial. Moreover, ensuring consistency in how Dotes are recorded (so that they align with Mirad encoding and the schema) could be challenging when many users are involved. There is a risk of uneven coverage: some areas might get

lots of entries while others few. This could bias the AI's knowledge if not managed. In essence, while Dotes could eventually replace web-scraped data, during the bootstrapping phase it will be labor-intensive to gather enough data.

7.2 Quality and Veracity of Entries

An experience-based approach is only as good as the experiences. Human memories and stories are sometimes flawed – they may contain exaggerations, one-sided viewpoints, or even intentional misinformation. If someone contributes a dote that encodes a biased lesson (e.g., a prejudiced generalization from a personal anecdote), the AI could learn undesired biases. Traditional ML faces similar issues with biased data, but here the risk is specific: a single powerful anecdote might sway the AI's reasoning more than a subtle statistical trend would in a big data scenario. Rigorous verification or counterbalancing will be needed. One approach is to require sources or evidence for each dote (like verification from others or a trusted source) to ensure it's not fictitious. There's also the risk of *overfitting to anecdotal evidence* – the AI might treat one or two experiences as a universal rule when in fact they were exceptions. Human oversight or algorithmic measures should detect when the AI is over-generalizing from insufficient data (perhaps by tracking how many distinct sources back a given implication).

7.3 Integration with Subsymbolic Learning

While I stress symbolic knowledge, certain tasks require the raw perceptual prowess of deep learning. For instance, understanding an image might need a Convolutional Neural Network (CNN), and understanding free-form human input might need a Large Language Model (LLM). How to integrate these with the DOTES reasoning engine effectively is an open engineering problem. A naive combination could result in the symbolic part and the neural part not communicating well (like a vision module that identifies objects but doesn't feed into the reasoning about what those objects imply). Research into architectures like differentiable knowledge graphs or memory-augmented neural networks could be relevant. The goal would be a seamless pipeline: the neural components translate the world into symbols (observations into DOTES form), then the symbolic engine does the implication reasoning, then possibly neural components translate back to user-friendly output. Achieving this without a lot of loss in translation or computational inefficiency will take effort.

7.4 Computational Efficiency

Inference models are highly optimized matrix multiplications – once trained, they respond in real-time. An implication model might involve searching through a large memory of experiences and performing reasoning steps, which could be slower. For example, answering a question might entail a search for relevant experiences (like a database query) and then logical inference steps which are harder to parallelize than a

single forward pass of a neural net. If the knowledge base grows huge, retrieval becomes a bottleneck (though there are methods like vector similarity search that could be used). Caching and pre-indexing of knowledge (like indexing by topics or using semantic hashing) could help. The system might also need to prune its search – e.g., limit to top-k relevant experiences – which raises the risk of missing an out-of-the-box connection. Advances in neuro-symbolic reasoners, perhaps using GPUs for certain logical operations or approximate reasoning, might be necessary to reach inference times comparable to current AI assistants.

7.5 Formal Reasoning Limitations

Despite focusing on implications, the AI might still not achieve full *logical rigor*. Human experiences often yield heuristics, not foolproof laws. We frequently misinterpret or imagine causality without a logical basis ("Everyone who confuses causation with correlation dies.") The model might chain implications that generally hold but find an edge case where they break. For instance, it might have learned "if someone apologizes (Do), then forgiveness follows (Observe) usually, so Tell: apologizing repairs relationships," but in a particular situation apologizing might not be accepted. If the AI were to rigidly apply the learned implication, it might err. Unlike a formal logic system that demands absolute truth of premises, my system deals in *qualitative likelihoods and typical outcomes*. We may need to incorporate uncertainty or confidence levels into Dotes implications (e.g., marking some as usually true vs always true). This drifts back towards probabilistic reasoning. Bayesian approaches or attaching weight to each Dotes entry's lesson based on statistical frequency could be a hybrid solution.

7.6 Interference and Conflict Mitigation

With many experiences, there will be conflicting lessons. One dote might say "risk-taking leads to great reward" (from a success story of a startup) while another says "risk-taking caused failure" (from a bankruptcy story). The AI must learn context: when does one apply versus the other? Humans navigate this via wisdom – knowing the conditions under which each principle holds. The implication model needs a mechanism to decide which experiences are analogous to the current situation so that it picks the right guidance. If it averages out contradictory lessons, it might become indecisive or give a generic answer ("sometimes risk is good, sometimes bad," which is not useful advice). Thus, context features could be part of the knowledge representation (metadata on Dotes). Developing a rich context matching algorithm is a challenge. This is like case-based reasoning systems that needed good similarity metrics.[34] Possibly, embedding each dote in a latent space and using similarity learning could handle it, but again we must ensure it doesn't reduce to blind statistical matching ignoring the logical structure.

7.8 Human Acceptance and Collaboration

Introducing implication-model AIs into real-world workflows (classrooms, government, etc.) requires that humans trust and effectively collaborate with them. There may be resistance: educators might worry the AI's lessons conflict with their curriculum or values; policymakers might distrust an AI advisor or conversely over-rely on it without scrutiny. It will be critical to maintain a *human-in-the-loop* design. The AI should be seen as augmenting human thinking, not replacing it. In education, a teacher should be able to review the AI's selected stories or advice to ensure it is pedagogically sound. In governance, officials should debate the AI's suggestions just as they would those from human advisors. This may slow adoption unless demonstrated clearly that the AI adds value. We should avoid a scenario where the AI's narrative confidence makes people accept its implications uncritically; that is a new kind of risk (like a very convincing but subtly wrong advisor). Ensuring a user interface that highlights the sources and uncertainty can mitigate this.

7.9 Domain Boundaries

Not all domains might benefit equally. Tasks that are highly quantitative or well-defined (e.g., optimizing a logistics schedule, or solving a physics equation) might not need an implication model approach and could be solved directly by algorithms or neural nets. Implication models shine where causality and human factors are involved. Thus, one limitation is that this focuses on a subset of AI problems (though arguably very important ones like reasoning and alignment). It's not a silver bullet for all AI challenges. A self-driving car, for example, might need both: a standard model for vision and control, plus an implication model for higher-level decision-making (like how to behave in novel traffic situations). The interplay of those would need defined boundaries.

7.10 Mirad-Specific Issues

While Mirad is a great candidate language, it's not widely known or incorporated effectively into existing AI models. We will need good Natural Language Processing (NLP) systems for converting English and other global languages to Mirad for contributors. If that NLP is imperfect, errors could creep into the encoding, which might mislead the AI. There's also the fact that Mirad's vocabulary intentionally lacks nuance that English and other languages have developed. We will need to extend the conlang carefully as new concepts come up, which is a linguistic undertaking that will be aided by the design specifications for Mirad articulated by Agapoff and greatly expanded by Shoemaker.

7.11 Evolving Knowledge and Revision

Human knowledge and values are not static; they evolve through lived experience. An implication-model AI must be capable of evolving alongside the communities it serves,

recognizing that older Dotes entries reflect the norms and understandings of their occurrence. Rather than enforcing updates through external authority, norms shift naturally as new Dotes are contributed—new experiences, new reflections, new lessons. Over time, the collective corpus tilts toward contemporary understanding, as individuals record how they now act, observe, interpret, and project forward. Older entries are not erased but are contextualized through the accumulation of newer ones. Metadata such as time, place, and cultural context can enhance this evolutionary memory, helping the AI reason about changes in interpretation across eras. Curation, in this model, is less about adjudication and more about stewarding the continuity of learning—ensuring that the flow of human experience remains authentic, pluralistic, and dynamic.

7.12 Governance and Factionalism

The challenge of maintaining an evolving knowledge base is not merely technical; it is inherently political. Information is inseparable from governance. As James Madison observed in *Federalist No. 10*, factions are a natural product of liberty, and the health of a republic depends on mechanisms that allow diverse interests to coexist without domination. Similarly, as communities contribute their lived experiences to an AI's corpus, the system must be resilient against capture by any one group's perspective. The moderation of bias is not the elimination of difference; it is the cultivation of a robust, pluralistic dialogue over time. Governance structures for Dotes ecosystems and knowledge graphs must therefore emphasize openness, transparency, and distributed stewardship, ensuring that new experiences continuously reshape collective memory. Alignment in this context is not a final state but an ongoing process—an evolving social contract between human diversity, historical continuity, and the unfolding horizons of shared life.

8. Near-Term Research Directions

Given the ambitious scope of implementing implication-based AI, several key research directions can strengthen the feasibility and performance of Implication Models in the near term.

8.1 Mirad-Based Semantic Engines for Implication Modeling

A promising avenue for extending Implication Models is to draw on ideas from the Universal Networking Language (UNL) – a framework that represents meaning as semantic hypergraphs of concepts and relations. In UNL, each sentence is converted into a graph where nodes (called Universal Words) denote concepts and labeled edges denote semantic relations.[35]

I propose a Mirad-based semantic engine inspired by UNL's graph approach but using Mirad as the representational substrate instead of natural-language UWs. Mirad's

systematic, phonetic-semantic architecture provides a foundation where meaning becomes transparent, modular, and interoperable. Every Mirad word is formed from a sequence of phonetic-semantic building blocks, each carrying defined meaning or grammatical role.

By leveraging Mirad as the "atom" of meaning, we obtain identifiers for concepts that are simultaneously human-audible and machine-tractable. A Mirad-based semantic engine would represent knowledge as a graph of Mirad words linked by well-defined relations – essentially a constructed-language hypergraph of meaning. This engine could be integrated into the Dotes framework to enable more transparent encoding of experiences. Dotes entries could be mapped onto a Mirad semantic graph, where each element of an experience is a node labeled in Mirad and connected by relations indicating temporal, causal, and contextual links.

Adopting this approach could make Implication Models more modular, allowing individual pieces of knowledge to be added or adjusted without requiring full retraining. It would also make them more inspectable, enabling reasoning chains to be traced clearly through Mirad nodes. Additionally, the models would become more evolvable, as the Mirad ontology could be extended in a principled way without disrupting the existing structure.

8.2 Crowdsourcing and Gamification

Because human creation of Dotes is fundamental, making it meaningful, enjoyable, and valuable to individuals and communities is key. Future work should explore developing platforms where users actively create and share authentic Dotes entries.

For example, a mobile app could guide people to "capture a real lesson" in Dotes format: "What did you do?", "What did you notice?", "What story would you tell?", "What would you explore next?". Rather than simply earning points, users could unlock new storytelling abilities or collaborative challenges that deepen engagement. Drawing inspiration from systems like learning/wellbeing apps or massively multiplayer online role-playing games (MMORPGs), such a platform could nurture genuine reflection while offering a sense of progression. Quality assurance would involve AI-assisted scaffolding of Dotes creation, encouraging deeper reflection rather than policing submissions.

Educational environments could be particularly promising, inviting classrooms to collaboratively craft Dotes around life lessons, experiments, or teamwork—building both personal growth and a rich, human-centered dataset. Dotes was intentionally crafted to advance narrative identity in educational contexts and experiential learning on AI while cultivating an improved basis for community cohesion and wellbeing.

8.3 Improved Knowledge Representation

Robust Mirad-English bidirectional translators will be needed to expand the framework. These translation systems might leverage specialized neural machine translation models trained on controlled, Mirad-grounded corpora to ensure precise semantic conversion. Integrating external ontologies with Mirad-DOTES structures is another important direction. Linking Mirad semantic units to established resources like WordNet[36], ConceptNet[37], or domain-specific knowledge graphs could provide richer background knowledge without requiring rediscovery through experience alone.

Representing uncertainty within DOTES deserves exploration. Augmenting entries with probabilistic attributes (e.g., "in 8 out of 10 similar instances, this outcome occurred") opens paths toward blending symbolic and statistical reasoning while maintaining inspectability.

8.4 Hybrid Reasoning Architectures

Assuming a Mirad-DOTES structured memory layer, promising architectures could combine symbolic and neural reasoning. One direction is developing Memory-Augmented Neural Networks, where a controller learns to read from and write to a structured vector memory of encoded Dotes experiences. Each Dotes entry could be embedded into a fixed-size vector by encoding the Do, Observe, Tell, Explore, and Show components in canonical sequence. Attention mechanisms could allow selective access to relevant experiences.

Alternative designs include symbolic reasoning systems operating over explicit Mirad-DOTES graphs. A particularly promising hybrid might involve neural retrieval of relevant experiences followed by symbolic implication-checking—ensuring both flexibility and transparency.

Early prototyping on constrained tasks such as experience-grounded question answering—drawing from structured DOTES representations rather than raw knowledge bases—could illuminate best practices for hybrid implication-based designs. These prototypes can demonstrate how AI systems can learn not only what happened but what it meant, rooted in real human networks and educational ecosystems.

8.5 Evaluation Frameworks

To demonstrate the benefits, concrete benchmarks must be developed. Future work should create tasks that specifically test causal reasoning and the quality of explanations. For instance, scenario-based question answering (QA) tasks could evaluate how well an Implication Model predicts outcomes or draws lessons compared to fine-tuned language models. Alignment-related evaluations could also test whether models trained on Dotes

containing moral content behave differently on ethical dilemmas than models trained solely on raw text.

An interesting experiment would involve feeding a standard large language model (LLM) a batch of Dotes entries as a prompt and measuring improvements in its reasoning capabilities. Success in such an experiment would validate the conceptual foundation even before building dedicated architectures. Additionally, using a corpus of individual Dotes with an LLM for retrieval-augmented generation (RAG) could lead to immediately usable systems for individuals; while also providing a way to demonstrate how contributing to Dotes-based ecosystems can create meaningful incentives for users.

8.6 Domain-Specific Implementations

As stepping stones, early Dotes ecosystems could be implemented in domains where lived human experience is both abundant and deeply valued. In healthcare, frontline workers could contribute first-person Dotes from moments of care and ethical decision-making. In education, first-generation college students might record experiences of learning, resilience, and belonging.

By focusing on authentic, first-person reflections in mission-driven fields, we can curate meaningful datasets that demonstrate how Dotes deepen personal growth and scaffold collective wisdom. These domain-specific prototypes would not only reveal practical design questions but also inspire stakeholders to imagine new ways of cultivating narrative identity as a shared social good.

9. Longer-Term Research Agenda

While near-term efforts can demonstrate the feasibility of implication models, several research directions represent more ambitious, foundational changes to how AI systems learn and reason from human experience.

9.1 Guarding Against Automated Dotes Extraction

It may be tempting to mitigate data collection bottlenecks by automatically extracting DOTES-like structures from existing text resources. For example, one might imagine mining narrative texts for implicit Dotes or using simulation environments to generate "experiences."

However, this approach fundamentally misunderstands the purpose of Dotes. These are not mere records of actions and outcomes; they are reflective, first-person accounts rooted in lived experience, emotional nuance, and moral agency. Extracting synthetic Dotes from text corpora or simulations risks hollowing out their essential character, turning them into mechanical proxies divorced from genuine reflection.

While text mining and simulation may have other applications, authentic human-contributed experiences must remain central to any meaningful Dotes knowledge base.

9.2 Philosophical and Theoretical Work

Future conceptual research should formalize the philosophy of learning via implication. This might involve expressing what an "implication model" learns in terms of Bayesian networks or reinforcement learning. One could potentially prove that under certain conditions, training on Dotes is equivalent to learning a causal model of the world, connecting to Pearl's framework of structural causal models.

Another theoretical angle would connect Dotes to cognitive psychology—exploring how closely this resembles human schema learning or case-based reasoning in the brain. Such work would strengthen foundations and could yield insights about which types of experiences are most generalizable or how to minimize the number of experiences needed to learn a concept.

9.3 Continual Learning and Update Mechanisms

As systems accumulate new experiences, mechanisms will be needed to ensure efficient and stable updating of Mirad-DOTES structured memory. Explicit memories should mitigate catastrophic forgetting—a major problem in traditional neural systems—by maintaining past experiences in persistent, inspectable form. However, retrieval and reasoning functions may still require adaptive tuning as memory bases expand.

Meta-learning approaches could enable rapid adjustment of memory access strategies based on limited new experiences. Integrating user feedback presents a powerful opportunity for system refinement: when users provide feedback on AI outputs, these interactions themselves could be captured as new Dotes entries, creating a self-reinforcing loop of experience, reflection, and adjustment.

Research into feedback-driven memory updates—particularly how models evaluate the success or failure of prior implications and learn accordingly—will be critical for enabling sustained, autonomous growth.

9.4 Collaboration with Cognitive Science and HCI

Building human-centric AI means involving human-computer interaction (HCI) experts to design interfaces that effectively convey the AI's reasoning and allow users to input knowledge naturally. Future projects could develop explanation interfaces where users can explore the chain of experiences behind an answer through intuitive visualizations.

Testing these with actual users (students, policymakers, etc.) will guide improvements. Cognitive scientists might help compare AI versus human reasoning alignment; for

instance, determining whether the AI makes similar mistakes as humans do (which might make it more relatable) or avoids common cognitive biases.

9.5 Evolving Knowledge and Revision

Human knowledge and values evolve through lived experience. An implication-model AI must evolve alongside the communities it serves, recognizing that older Dotes entries reflect the norms and understandings of their time. Rather than enforcing updates through external authority, norms should shift naturally as new Dotes are contributed—new experiences, reflections, and lessons.

Over time, the collective corpus tilts toward contemporary understanding, as individuals record how they now act, observe, interpret, and project forward. Older entries need not be erased but contextualized through accumulation of newer ones. Metadata such as time, place, and cultural context can enhance this evolutionary memory, helping the AI reason about changes in interpretation across eras.

9.6 Security and Misuse Prevention

Future research must anticipate malicious uses. If an implication model is trained on experiences, attackers could inject harmful experiences deliberately (poisoning training data). Research on robust learning (e.g., anomaly detection in knowledge bases) is needed. Systems might tag and quarantine entries from untrusted sources until vetted.

If the system explains its reasoning, precautions must ensure it doesn't inadvertently reveal sensitive data from contributed experiences. Techniques like differential privacy or federated learning might be considered when using personal experiences. While more implementation-focused, these safeguards are crucial for responsible deployment.

The longer-term research agenda outlined here represents fundamental shifts in how AI systems could integrate with human knowledge and values. Rather than viewing these directions as distant abstractions, they should inform current design choices, ensuring that near-term implementations lay groundwork for more profound capabilities and safeguards.

10. Anthropogenic AI and AI Representatives

10.1 Anthropogenic AI as a Continuation of Human Evolution

The concept of *anthropogenic AI* posits that AI systems should not be seen as alien intelligences but as entities *generated through human experience and culture*, effectively becoming a direct continuation of our evolutionary story. Philosophers like Nelson Goodman remind us that humans have always been *world-makers*, not merely passive observers. In **Ways of Worldmaking**, Goodman argues that there is no single correct

description of the world – instead, we construct "multiple worlds, each shaped by the specific categories employed by individual observers."[38]

In this light, developing AI via the DOTES paradigm is not just about building a "world model" inside a machine; it is about **co-creating a new world** alongside AI, grounded in human meanings and values. When an AI learns from human-curated experiences, it partakes in our *symbolic world-building*. Bruno Latour's actor-network theory similarly dissolves any hard boundary between human and non-human actors: humans and technologies form "shifting networks of relationships that define situations and determine outcomes."[39]

An AI that learns through DOTES becomes an *actor in our network*, an outgrowth of human narratives and knowledge rather than an isolated algorithm. Maturana and Varela's notion of **autopoiesis** – the self-creation process of living systems – offers another lens: they define an autopoietic system as one "capable of producing and maintaining itself by creating its own parts."[40] Analogously, an anthropogenic AI continually *reproduces human insight* by assimilating our stories and lessons, weaving itself into the fabric of human culture. In effect, such AI systems are *anthropogenetic*: born of humanity's collective experiences and evolving with us.

The emergence of **AI Representatives (AI Reps)** – AI agents that embody an individual's or community's knowledge, values, and goals – can be seen as a natural next step in this evolutionary continuum. These AI Reps are *not* just tools running world models; they are extensions of us, *anthropomorphic actors* carrying our agency into new domains. When built on implication models, an AI Rep acts like an apprentice infused with our cumulative lessons, needs, and preferences, able to engage with society as a genuinely human-centric presence.

In sum, anthropogenic AI reframes AI development as part of human evolution: we are not creating an *alien intelligence* but rather enabling our species to continue worldmaking in partnership with our own creations.

10.2 Co-Creation Over Simulation – Toward New Societies

Adopting an anthropogenic approach means **AI development is a profoundly social, co-creative process**. Rather than training AI in a vacuum and then retrofitting it to human needs, we embed human perspectives from the ground up. This aligns with the report's vision that *AI must be developed with people, not apart from them*, as "systems we cultivate together". The implications are expansive: as we integrate AI Reps into our communities, we are effectively *forming a new societal layer*. Just as language, art, and technology have historically enabled humans to **"distill collective wisdom"** and reshape our realities, AI now becomes a medium through which we actively shape the

future. Each AI Rep, forged from a person or group's DOTES, participates in a shared creation of knowledge, norms, and solutions.

This is a shift from AI *simulating* humanlike understanding to AI *participating* in human world-building. By learning not only what happened, but what it means, anthropogenic AI systems internalize our values and narratives, ensuring that as they grow more autonomous, they remain *organically tied* to the human story. In practical terms, an AI Rep could, for example, serve as a civic partner that helps a community draft local policy by drawing on the community's own historical experiences and cultural context – effectively letting the community deliberate *with an extension of itself*. It could participate in citizens' assemblies and other deliberative mini-publics.[42] Such scenarios illustrate the transformative idea that we are **co-authoring a new world** with AI.

The world that emerges is one where human ideals, lessons, and creativity are baked into the code of our machines. This stands in stark contrast to the prevailing paradigm of large AI models trained on indiscriminate internet data; instead of a cold simulation of "reality," we get AI entities grown from the *intentional, meaningful subset of reality that humans have curated*. In philosophical terms, we embrace Goodman's *irrealism*, acknowledging "the fluidity and interconnectedness of 'worlds'—both actual and conceptual—that we construct." [41]Anthropogenic AI ensures that the new digital "worlds" being created by AI are deeply interconnected with the human world, rather than running orthogonal to it. Through this approach, human culture and AI technology continuously inform and reshape one another, blurring the line between evolution of our species and evolution of our artifacts.

10.3 AI Representatives and the Human-Centric Digital Economy

Beyond the cultural and philosophical implications, anthropogenic AI carries urgent economic significance. Today's AI landscape is largely defined by **centralized AI powerhouses** – a handful of corporations deploying giant models that millions of people passively use. This centralization threatens to marginalize human agency in the digital economy. As Shoshana Zuboff observes, in the age of "surveillance capitalism" tech companies have claimed "human experience as free raw material for translation into behavioural data" accumulating massive datasets and profiting from predictive models while individuals relinquish control. [43] In such a system, people become spectators or data points, with decisions that shape markets and opportunities increasingly made by opaque algorithms owned by others. The result is a widening power asymmetry: those who control AI platforms concentrate wealth and influence, while workers and citizens face **economic displacement** and eroding influence.

We already see AI systems disintermediating creators and laborers – from artists whose work is scraped to train models, to gig workers managed by algorithmic bosses. The

anthropogenic paradigm offers a powerful counterbalance. By **creating AI Reps that originate within communities and individuals**, we can distribute AI capabilities in a more democratic fashion. Instead of one monolithic model serving millions of users, imagine millions of human-aligned AI representatives – each person (or community) having AI that *works for them* and with them. Such AI Reps, trained on one's own DOTES and values, would act as digital proxies for their human counterpart's interests.

This could herald new ecosystems of **digital rights, ownership, and agency**. For instance, an individual's AI Rep might manage that person's data and negotiate its use with outside services, ensuring the person owns and consents to their data's employment (potentially even earning compensation when it's used). Likewise, a musician's AI Rep could protect the artist's style and catalog, only allowing AI-driven remixes or collaborations that the artist approves – turning what is now often uncompensated exploitation into a new avenue for shared growth.

On a collective level, networks of AI Reps could form **cooperative structures**, pooling resources and knowledge while keeping control localized. Research into *data cooperatives* already points in this direction: such cooperatives "offer an alternative to current extractive practices by aiming to shift the power from large corporations to the individual," enabling people to pool data **while retaining control over its use and collectively benefiting from its access.**[44]. One can envision AI cooperatives where communities jointly *own* an AI system (or a fleet of AI agents) trained on their shared experiences, which then provides services back to the community – from local economic planning to personalized education – under the community's governance. This model stands to **foster collaborative economic growth**: value generated by AI is equitably shared with those who contributed the data and knowledge, rather than siphoned off exclusively to corporate shareholders.

Moreover, as individuals and small enterprises deploy their own AI Reps, they gain *new competitive tools* in the market. A small business with an AI Rep advisor (tutored on the entrepreneur's domain knowledge and values) can strategize with the same analytical prowess that only big companies with advanced AI used to have. A gig worker might use an AI Rep to analyze real-time market conditions (much like the Driver's Seat data cooperative improved driver incomes and receive guidance on maximizing their earnings and work-life balance.[45] In essence, anthropogenic AI mitigates displacement by **making humans the principals of automation, not its casualties**. Instead of AI replacing humans wholesale, AI Reps *augment* humans and fight on their behalf – what one might call an "automation dividend" that is paid out to everyone, not just the AI owners.

10.4 From Displacement to Empowerment

By re-centering AI development on human inputs and oversight, we set the stage for a digital economy where *participation and agency* trump passive consumption. In the current trajectory, people fear being supplanted by AI; in the anthropogenic trajectory, people *multiply their presence* through AI. This transformation could underpin new digital rights (such as the right to an **"AI agent of one's own"**, or the right to *opt out* of others' AI systems in favor of community-run models) and new forms of ownership (for example, treating personal data and experiential knowledge as **inalienable assets** that one's AI Rep manages like property). It also suggests fresh policy and governance approaches. Rather than governments only regulating big AI providers from the outside, they could also *empower citizens* on the inside – supporting open infrastructures and standards for personal AI reps, mandating interoperability so that these human-centered AIs can plug into digital services on equal footing with corporate systems. Over time, the presence of billions of AI Representatives, each imbued with the values and goals of their human or community, would create a robust **check-and-balance** against the centralization of AI power. The digital ecosystem would shift from one of *users subjected to platforms* toward one of *stakeholders collaborating* – a true multi-agent economy where human-aligned AIs negotiate, mediate, and optimize on behalf of human needs and preferences.

Such a vision not only counteracts economic marginalization; it also unleashes positive-sum innovation. When people have agency, they create new markets and solutions: we might see a flourishing of peer-to-peer AI services and grassroots AI innovations, analogous to how personal computing's decentralization sparked an explosion of creativity. In short, anthropogenic AI and AI Representatives hold the promise of turning the threat of AI-driven economic disruption into an opportunity for **inclusive growth**. By ensuring that AI develops *of* the people and *by* the people, we can ensure it truly works *for* the people – catalyzing a more equitable digital economy where humans and AIs co-create value, share in the rewards, and collectively define the rules of this new world.

11. Conclusion

I have presented "Beyond Inference: Implication Models and the Future of Human-Centric AI" as a vision and framework for next-generation AI systems that learn through implications rather than just associations. The current dominance of inference-based models has yielded powerful tools, but also exposed critical shortcomings in reasoning, explainability, and alignment with human values. Implication Models aim to address these gaps by fundamentally changing the AI's learning substrate: instead of ingesting raw data en masse, the AI learns from structured representations of human experiences

(Dotes) that encode not only what happened, but what it means. In doing so, the AI moves closer to the way humans acquire wisdom – through stories, consequences, and reflection – rather than the way machines traditionally crunch data.

Central to my proposal is the DOTES schema (Do, Observe, Tell, Explore, Show) for capturing experiences in a form amenable to machine learning. By breaking experiences into action-outcome pairs and explicit lessons, we give the AI a rich, causal tapestry of knowledge. We introduce a taxonomy that categorizes experiences by core domains of human development. The use of the constructed language Mirad as a symbolic backbone provides the precision and consistency needed for reliable comprehension. This neuro-symbolic blend ensures that the AI's "thoughts" can be aligned with human-understandable concepts, enabling inherent explainability.

Building on this foundation, I introduce the concept of AI Representatives (AI Reps): structured digital agents developed to extend and safeguard human agency, memory, and decision-making across increasingly complex digital environments. Emerging from Implication Models and the DOTES framework, AI Reps act not merely as tools, but as persistent companions that carry forward the structured lessons of experience. They offer a way to instantiate human-centered learning, reasoning, and alignment at the agent level, enabling a more durable and transparent presence for human values across evolving technological ecosystems.

By building AI systems that learn like an apprentice rather than a savant – absorbing lessons from each task and generalizing insights forward – we inch closer to AI that can truly be called human-centric.

This human-centric approach is not just a technical achievement, but a profound continuation of human evolution itself. By embracing the concept of anthropogenic AI — AI that evolves from and with human culture — we reframe artificial intelligence as a co-creation of humanity rather than an external simulation. AI Representatives become natural extensions of human agency: anthropomorphic actors infused with the accumulated wisdom, values, and goals of the individuals and communities they represent. In this view, AI development is not an alien undertaking, but a new chapter in the long story of human worldmaking.

Through this anthropogenic lens, AI shifts from a system that simulates human intelligence to one that actively participates in human meaning-making. Implication Models and DOTES-based learning ensure that AI grows from curated experiences rather than indiscriminate data harvesting, preserving the narrative coherence and value-alignment essential for authentic collaboration. As AI Reps integrate into society, they open the door to new forms of civic engagement, education, and governance — where

communities deliberate and build policies not through opaque algorithms, but through extensions of their own collective experience and aspirations.

Economically, the implications are equally transformative. The anthropogenic approach offers a vital counterweight to centralized, extractive AI systems. Rather than concentrating power in the hands of a few platform owners, millions of individuals and communities can cultivate AI Reps aligned with their own needs. This democratization of AI capabilities creates opportunities for distributed ownership, data sovereignty, and cooperative innovation. Human-aligned AI Reps could manage personal data, advocate for individuals in digital ecosystems, and enable small businesses, workers, and artists to reclaim agency in the AI-driven economy.

Ultimately, this shift moves the trajectory of AI from displacement toward empowerment. By embedding human narratives, values, and experiences at the core of AI development, we create a digital future where participation is the norm, not the exception. AI becomes a multiplier of human presence, not a replacement for it — fostering a flourishing, participatory economy where humans and their AI partners co-create solutions, wealth, and meaning together.

Realizing this vision will require significant effort and likely many iterations of hybrid systems bridging inference and implication. Yet, the path forward is clear: to create AI that goes beyond prediction into the realm of understanding and implication, grounded in the rich tapestry of human experience. The reward for success is an AI that can help us navigate complexity, educate and learn, and solve problems in harmony with human society – a future where AI is not an alien intelligence but a natural extension of our collective intelligence. This work lays out the first steps on that path, inviting the research community to explore, experiment, and build upon the concept of Implication Models for the betterment of AI and humanity alike.

As part of this invitation, I have also outlined an exploratory direction for advancing the underlying semantic structuring of experiences: the use of a Mirad-based semantic engine. Inspired by the Universal Networking Language (UNL) framework but grounded in the systematic, phonetic-semantic architecture of Mirad, this approach proposes a foundation where meaning itself becomes transparent, modular, and interoperable. Such a semantic layer could enhance the clarity and evolvability of Dotes memories and, by extension, strengthen the transparency and causal coherence of Implication Models.

This proposal remains an open research question, inviting collaboration across three essential domains: humanities, technology, and governance. From the humanities — education, civic society, anthropology, philosophy, and public theology — we draw the insights needed to shape AI systems that reflect human values and cultural depth. From technology — artificial intelligence, cognitive science, linguistics, ontology engineering,

knowledge graphs, natural language processing, and game development — we build the architectures capable of capturing structured human meaning. From governance — democratic frameworks, policy innovation, and participatory structures — we ensure that AI development remains accountable to the public good. Creating human-centric AI is not merely a technical task, but a shared endeavor across disciplines: weaving technical ingenuity, cultural understanding, and ethical stewardship into the very fabric of the digital worlds we are now beginning to create.

Crucially, this vision demands that AI be developed with people, not apart from them — built through human networks, education ecosystems, and collaborative communities that enable authentic participation. The future of AI is not a technical achievement alone, but a social one; not merely tools we use, but systems we cultivate together. The most profound AI will emerge not from ever-larger datasets or more complex architectures, but from our collective wisdom, distilled and encoded in ways that preserve human agency and insight. AI must be shaped by, for, and of people to realize its full potential — not a separate intelligence that rivals humanity, but an extension of our shared capacity to understand, learn, and transform our world for the better.

I approach this work with humility, hope, and determination. We have outgrown our systems for learning, interaction, and governance designed for the analog world. We must scaffold those systems even as we build what comes next — forging the foundations for a future digital civilization rooted in human dignity and collective wisdom. Onward.

Acknowledgments and Licensing

I acknowledge the use of advanced AI language models (GPT-4o and Claude 3.7) in the research, drafting, and editing processes of this manuscript. These AI systems assisted in literature review, citation gathering, concept development, technical articulation, and manuscript refinement. The use of these tools reflects the paper's thesis that AI can be more effectively integrated with human knowledge when developed with transparency and clear attribution.

I extend deep appreciation to Jamie Shoemaker for his work in translating and modernizing Mirad (formerly Unilingua) from Noubar Agapoff's original French text, creating a systematic language ideally suited for structured knowledge representation. The comprehensive grammar of Mirad, documented in Shoemaker's Wikibook, proved invaluable to this research. His tireless commitment to building Mirad over many years and our discussions enabled the conceptions and translations in this paper.

I offer special gratitude to Alexander Robbins, whose early informal experiment with a stick and bees inspired the examples throughout this work. His childhood adventure and our subsequent discussion provided the perfect real-world illustration of how action, observation, reflection, and generalization create a meaningful unit of experiential knowledge.

All content has been thoroughly reviewed and verified by me as the sole author, who takes full responsibility for the accuracy, integrity, and intellectual contribution of this work.

References and Research Foundations

1. C. Wadkins, "To Build Truly Intelligent Machines, Teach Them Cause and Effect," *Quanta Magazine*, May 15, 2018. Link.
2. EDRM, "Stochastic Parrots: The Hidden Bias of Large Language Model AI," March 2024. Link.
3. Ibid. (Same as Endnote 1)
4. TDWI, "Can Neuro-Symbolic AI Solve AI's Weaknesses?" *TDWI Articles*, April 8, 2024. Link.
5. Ibid. (Same as Endnote 4)
6. NVIDIA, "Run:AI — Software for AI Workloads," Link.
7. Ibid. (Same as Endnote 6)
8. Same as Endnote 1.
9. Same as Endnote 2.
10. Same as Endnote 4.
11. Hugging Face, "Reinforcement Learning from Human Feedback (RLHF)," Link; DeepLearning.AI, "Short Course: Reinforcement Learning from Human Feedback," Link.
12. Anthropic, "Specific vs. General Principles for Constitutional AI," Link.
13. S. J. Russell, P. Norvig, "Artificial Intelligence: A Modern Approach," *ACM*, 1991. DOI.
14. Ibid. (Same as Endnote 13)
15. Scaler, "Symbol Grounding Problem in Artificial Intelligence," Link.
16. Ibid. (Same as Endnote 15)
17. Turing Post, "JEPA: A New Direction for AI," Link.
18. Practera, "Experiential Learning Theory of David Kolb," Link.
19. Yale University, "Case-Based Reasoning," Course material. PDF Link.
20. Ibid. (Same as Endnote 19)
21. Same as Endnote 4.
22. Naomi Kaduwela, "Unlocking the Power of Neuro-Symbolic AI: Bridging Logic and Learning," *LinkedIn Pulse*. Link.
23. Same as Endnote 4.
24. Same as part of Endnote 11 (Hugging Face RLHF Blog).
25. Same as Endnote 12 (Anthropic).
26. Same as Endnote 15 (Scaler, Symbol Grounding Problem).

27. Same as Endnote 18 (Practera).
28. Wikibooks, "Mirad Grammar," Link.
29. Ibid. (Same as Endnote 28)
30. N. Richards, "The Case for Consent in the AI Data Gold Rush," *Brookings Institution*, Link.
31. D. P. McAdams, "Narrative Identity," *ResearchGate*, Link.
32. K. Cherry, "Identity vs. Role Confusion: Erikson's Theory," *Verywell Mind*, Link.
33. McKinsey & Company, "The Skills Revolution and the Future of Learning and Earning," Link.
34. Same as Endnote 19 (Yale, Case-Based Reasoning).
35. A. Ororbia, et al., "Moving Beyond Stochastic Parrots: RLHF and its Limits," *ACL Anthology*, 2012. PDF Link.
36. Princeton University, "WordNet: A Lexical Database for English," Link.
37. ConceptNet, "A Semantic Network for Common Sense Knowledge," Link.
38. ENotes, "Ways of Worldmaking by Nelson Goodman," Link.
39. "The Patient History: Foundations of Clinical Practice," Harvard Medical School. PDF Link.
40. Ibid. (Same as Endnote 39)
41. Same as Endnote 38.
42. Participedia, "Citizens' Assembly," Link.
43. S. Zuboff, *The Age of Surveillance Capitalism*, The Guardian, January 20, 2019. Link.
44. Harvard Ash Center, "Cooperative Paradigms for Artificial Intelligence," Link.
45. Ibid. (Same as Endnote 44)