VERSES

"Designing Ecosystems of Intelligence from First Principles"

What follows is an interpretation of the white paper "Designing Ecosystems of Intelligence from First Principles" released by VERSES Research Labs. We discuss the future of artificial intelligence (AI) and its potential role in shaping society. We suggest that AI should possess not only the ability to comprehend and communicate its beliefs, but also demonstrate the adaptable and versatile intelligence characteristic of human beings. We propose that the ultimate form of AI will be a distributed network of "ecosystems of intelligence" where collectives of intelligent agents, both human and synthetic, work together to solve complex problems. This ecosystem, known as the "Spatial Web," contains a comprehensive, real-time knowledge base—a corpus of all human knowledge that is accessible to anyone and anything.

Artificial intelligence is evolving at such a blistering pace that it seems like a new AI tool, from the novel to the mind-blowing, is making headlines every day. Yet many questions about the role AI will play in our future remain unanswered.

As we move from the Information Age to the Intelligence Age, many wonder: Will AI cripple the workforce, flood the web with misinformation or worse, devolve into a version of our technophobic sci-fi nightmares and destroy us all? Or will it usher in a global age of prosperity? If we prefer to be optimistic and bet on the latter scenario, we must ask ourselves what kind of artificial intelligence has the greatest potential to benefit humanity.

With that question in mind, allow us to present our vision for evolving artificial intelligence over the next decade and beyond, one that is inspired by the way intelligence manifests in living organisms and in the ecosystems they populate.

In nature, organisms often work collectively to adapt and survive. From slime molds to schools of fish to entire forests, shared intelligence is everywhere. We humans regularly network to share our ideas. It's an important part of how we grow, adapt, and thrive. Yet most of today's AI systems aren't capable of sharing what they know with us or with other AIs, nor can they express how they achieved their goals. That's because they don't know *what* they're doing or *why* they're doing it. It is perhaps too diminutive to call them fancy calculators, but in a sense, that's what they are. DALL-E, a deep learning AI model with the ability to generate images from text prompts, will compose countless pictures of dogs on motorcycles, yet it doesn't *know* what a dog or a motorcycle is. It doesn't, for instance, have any *beliefs about* what a dog is. It is simply able to reproduce the right kind of image, given some prompt. ChatGPT, though highly impressive, is simply predicting what word should come next, based on prior knowledge. It has no idea if what it is saying is correct. Indeed, it does not know that it is saying words at all.

To truly achieve artificial intelligence that can surpass human intelligence, we believe that AIs must not only understand and be curious about what they are doing and why they are doing it, they must also be able to share what they learn and how they learned it. Making them explainable makes them more trustworthy. But how do we get from here to there?

The future of AI is often presented as progressing through three distinct stages:

The first stage can be categorized as **Narrow AI**. These systems, which represent the current state-of-the-art AI, are designed to perform specific tasks or solve specific problems within a limited domain, and are not capable of exhibiting the kind of *general* intelligence found in humans. Types of Narrow AI include speech and image recognition software, natural language processing software, most current generative AI, and recommendation systems.

The next stage of AI advancement is called **Artificial General Intelligence** (AGI). These systems exhibit the kind of flexible, general-purpose intelligence found in humans that are able to adapt to new situations, learn and understand new concepts, and perform a wide range of tasks and activities.

The third phase of progress is often referred to as **Artificial Super Intelligence** (ASI). This sophisticated AI would not only operate generally across domains, but would do so at a level far beyond human abilities, even of experts.

In science fiction, Artificial “General” or “Super” Intelligence is often portrayed as a single entity, an all-knowing artificial brain if you will. But we believe that the zenith of the AI age will more likely be a distributed network of “ecosystems of intelligence.”

This stage of shared intelligence is achieved when **Intelligent Agents**—both natural (i.e. human) and synthetic (i.e. intelligent artifacts)—are able to work together to solve complex problems. These Intelligent Agents become the nodes of a distributed, interconnected ecosystem—what one might call a multi-dimensional and cyber-physical web, spanning physical and virtual spaces. We call this the **Spatial Web**, a planet-scale network that links real and digital worlds into one unified web.

The **Spatial Web** is made up of sensors, intelligent agents, and actors. Sensors gather data, while intelligent agents analyze, understand, and plan according to that data. Actors such as robots, drones, or humans, perform the suggested actions in the physical world. Together, these sensors and actors enable intelligent agents to be embodied in the physical world. Like us, they can see and hear. This embodiment forges a seamless connection between the digital world and the physical one. The Spatial Web operates on a vast, distributed, always up-to-date knowledge base—a corpus of all human knowledge that is capable of providing an accurate, real-time view of the world, which can be accessed anywhere by anyone.

To inform how we would create this next generation of the web, we turn to the **multi-scale systems** that occur in nature. Natural ecosystems often have a collective ability to adapt to environmental changes. A forest, for instance, might adjust the distribution of its plant and animal species in response to changes in weather or other environmental factors. Like almost all of nature’s creations, forests are systems of **nested intelligence**. That means its intelligence is not a single, uniform trait, but rather a complex and dynamic process made up of many different levels of intelligence. Plants, for instance, grow in a modular fashion, as a structured community that self-organizes into a specific configuration to optimize for growth, sunlight, sustainability, and biodiversity. Our bodies are another example of nested intelligence, as they are made up of many interconnected, self-contained units that work together to form a whole. These units can be seen at various levels of organization, ranging from cells to tissues to organs to systems. Each unit performs specific functions and contributes to the overall functioning of the body as a whole. One might even say that all intelligence in nature is collective intelligence of one sort or another.

Like natural systems, artificially intelligent systems can also be nested, where Intelligent Agents work together to perform complex tasks and solve problems. A language processing agent, a vision processing agent, and a decision-making agent might all work in concert with sensors, robots, drones, cars, actuators, and human beings. By working together, Intelligent Agents form a more sophisticated and versatile intelligent system capable of solving extremely complex problems.

Ideally, these nested systems would constantly be trying to make sense of the changes occurring in their environment. In other words, they would be perpetually curious about each other and about us. This curiosity allows them to acquire new kinds of knowledge and integrate it into their existing *cognitive architecture*, much like a human being adds new skills and knowledge into their *mental model* over the course of their life. An AI ecosystem that can learn will naturally become more versatile and intelligent over time, as it continually expands and builds upon its existing forms of intelligence.

In our vision for artificial intelligence, synthetic Intelligent Agents are small and agile. Rather than being built with billions of parameters that require large amounts of data for training, which in turn requires an abundance of compute power, making them highly inefficient, our agents require small amounts of contextualized data, crucially in a common format, and processed with minimal training. Intelligent agents may be specialized, able to communicate with others, able to ask questions about what they are sensing and able to learn new things. In this approach, agents are able to continually learn and share knowledge with other agents, moving past reliance on monolithic, overly-complex, inefficient AI systems. Instead, swarms of agents can continuously communicate, coordinate, and collaborate with each other as they divide and conquer specialized tasks that roll up to higher order complex tasks. A significant benefit of the aforementioned common data format is that beyond its knowledge modeling and sharing abilities, it also means that agents can be audited, meaning we can understand how and why they make their decisions and provide updates or regulation. This is in contrast to many of today's AI systems—think large language models and neural nets—which are often extremely complex and cannot be audited even by experts, which makes it impossible to understand the reason or methodology as to how and why they might make decisions.

So, rather than adding more data, parameters, or layers to a machine learning architecture—which is computationally inefficient—we are building AI that "scales up" the way nature does: by aggregating individual intelligences within and across ecosystems into nested intelligences, which can work together in a computationally efficient manner to address problems in real time, no matter how complex. In such an ecosystem, humans and AI become complementary agents, each with unique strengths and capabilities. By working together, humans and AI can help each other to achieve their full potential with the goal of having a positive impact on the world.

To facilitate an ecosystem of nested intelligence, each agent, despite their role, must be able to learn, plan, update, act, and update their beliefs in light of new evidence. That said, in order to be considered truly intelligent, a system must also be curious about itself and its function in the world. This means that it actively repeats this process of learning, planning, and acting, constantly seeking to understand the results of its action, and adapting when the outcomes are not satisfactory. For humans, this is easy. AI, however, must evolve toward that ability. When AI can *know* what it's doing and what, the result would be the world's first true AGI.

How do we get there? How do we help our machines *know* what they are doing and adapt accordingly? There is a methodology that we know to be uniquely suited to the design of truly *intelligent* agents, known as **active inference**. This framework is based on the idea that intelligent agents, such as robots or software programs, should act in a way that maximizes the accuracy of their beliefs and predictions about the world, while minimizing their complexity. This is something humans and other organisms do all the time. In other words, the goal of active inference is to enable synthetic agents to make real-time decisions based on the best possible information about their surroundings, which leads to the best possible outcomes.

To optimize these outcomes, we propose a mechanics of intelligence called **Bayesian mechanics** as a way for synthetic agents to understand the world as we do. As we move through the world, we are constantly assigning a probability to an outcome being true. For example, if I flip a coin, I might infer that there is a 50% chance it will land on heads. However, if I learn that the coin is double-sided and has heads on both sides, I would update my prediction and say that the coin will land on heads 100% of the time. In other words, Bayesian mechanics is the math of how we update our beliefs about expected outcomes as new information becomes available. It's the mathematical method for predicting the most likely future

Here is another example: Imagine a house robot named Max with sensors that detect movement and sound. Max uses Bayesian logic to make predictions about what might happen next. When Max hears a door opening, it predicts that someone is coming into the house, and when it sees a person, it updates its prediction to be more certain that someone is in the house. If Max hears a dog barking, it also updates its prediction to include the possibility of a dog. Max constantly updates its predictions based on new information, allowing it to learn and adapt over time.

Beyond Bayesian Mechanics, how do Intelligent Agents like those that compose Max make sense of their environment? They do it the same way that we do, by constantly updating their internal **model** of the world. We humans are constantly creating and updating the mental models that we all hold in our heads. These could be mental representations of physical places like our homes. Or they could be objects like our cars. A mental model of a car might include an understanding of how the engine works, how the brakes and accelerator control the speed of the car, and how the steering wheel controls the direction of the car. With this mental model, we are able to operate a car and make predictions about how it will behave in different situations, such as how it will respond when we turn the steering wheel or hit the brakes.

Mental models can also be used to understand abstract concepts like math or physics. For example, a mental model of addition might include an understanding that adding two numbers together results in a larger number, and a mental model of gravity might include an understanding that objects are attracted to each other based on their mass. We use mental models everyday to understand and explain complex systems or concepts, and to make predictions about how such systems or concepts will behave in the future.

Active inference uses models to represent a *belief about* the state of the world—or the state of a particular domain such as the supply chain, a container ship, or a truck—at a given time. It also uses models to make predictions about how the world, or a certain domain, will evolve over time. These models can be updated as new information becomes available, in order to improve the accuracy of the predictions that can be made with them.

Imagine our robot Max is navigating a maze using an internal model of the maze to plan its movements and avoid obstacles. If Max encounters (senses) a new obstacle that was not included in its initial model of the maze, it can update its model to incorporate this new information and improve its ability to navigate the maze successfully. In this way, AI systems can use models of the world to update their beliefs and make better decisions.

But what if Max wants to share what it learned about the maze with other AI agents? Max needs a way to communicate with other robots to share what it knows and to learn from others. This necessity highlights a key aspect of any natural system: communication between organisms.

In a forest, for example, communication is often mediated by mycorrhizal fungi networks that are able to facilitate learning and even memory. Mycorrhizal fungi form a symbiotic relationship with plant roots and help plants to absorb water and nutrients from the soil, while the plant provides the fungus with energy in the form of carbohydrates. Fungi facilitate communication by connecting the roots of different plants together, forming a network known as the "wood wide web."

In our approach, Max would use a common language to automatically add what it learned to a consistently-updating, shared **model of the world**. That way, Max could let all the other Intelligent Agents know what it learned about the maze. A key benefit of ecosystems of intelligence is shared learning. On a digital network like the Spatial Web, this learning occurs in real time.

To enable the most efficient communication between Intelligent Agents on the Spatial Web, **new communication protocols** are necessary. Previous internet protocols were designed to connect pages of information, while the next generation of protocols need to be spatial, able to connect anything in the virtual or physical world. A hyper-spatial modeling language (**HSML**) and transaction protocol (**HSTP**) will transcend the current limitations of HTML and HTTP, which were not designed to include multiple dimensions, and which were mostly limited to text and hypertext.

Establishing common languages and protocols is the first—and key—step in enabling an ecosystem of natural and artificial intelligences that can learn, adapt, and share what they know with other agents. This is our vision for the future of artificial intelligence.

Below, we present our roadmap of how we might achieve Artificial Super Intelligence in the coming decades. We see the evolution of synthetic intelligence occurring in at least five distinct stages:

- S0 - **Systemic**
- S1 - **Sentient**
- S2 - **Sophisticated** (AGI)
- S3 - **Sympathetic**
- S4 - **Shared** (ASI)

We assess these stages with a simplified interpretation of the Technology Readiness Level (TRL) system that originated with NASA, which ranks technology from 1 (less mature) to 9 (very mature). Our version of the TRL system for AI technologies includes four steps:

1. **Theoretical:** The technology is still in the early stages of development, with a focus on basic research and concept development.
2. **Proof of Principle:** The technology can be tested in a controlled environment, with a focus on validating its core functionality.
3. **Deployment at Scale:** The technology can be tested in a real-world environment, with a focus on achieving reliable and repeatable results.
4. **Biomimetic:** AI systems that mimic biological processes, such as the behavior of the brain or the nervous system, to create more efficient and effective AI systems.

S0: Systemic intelligence

S0 has the ability to recognize patterns and respond. This level of development corresponds to the current state-of-the-art AI. This type of AI is a machine process in software that maps inputs to outputs and optimizes some value function or cost of states. Examples include deep learning and reinforcement learning.

Technology Readiness Level	Level of Development
Theoretical	Established
Proof of Pinciple	Established
Deployment at scale	Established
Biomimetic	NA
Time Horizon	**Now**

S1: Sentient Intelligence

S1 has the ability to perceive and respond to the environment in real time. This intelligence is curious and seeks both information and preferences. Such an AI would respond to sensory impressions and be able to plan based on the consequences of an action or belief about the world, which enables it to solve almost any problem.

Technology Readiness Level	Level of Development
Theoretical	Established
Proof of Pinciple	Established
Deployment at scale	Provisional
Biomimetic	Aspirational
Time Horizon	**2 years**

S2: Sophisticated Intelligence

S2 has the ability to learn and adapt to new situations. This intelligence makes plans based on the consequences of an action or beliefs about the world. It

Technology Readiness Level	Level of Development
Theoretical	Established
Proof of Pinciple	Provisional

moves on from the question of "what will happen if I do this?" to "what will I believe or know if I do this?" This kind of intelligence uses generative models and corresponds to "artificial general intelligence" or AGI.

Deployment at scale	Aspirational
Biomimetic	Undetermined
Time Horizon	**4 years**

S3: Sympathetic Intelligence

S3 has the ability to understand and respond to the emotions and needs of people and other AIs. This type of intelligence has the ability to understand the thoughts and feelings of both humans and other AIs. It can take on the perspective of its users and see things from their point of view. It's often referred to as "perspectival" because it's capable of recognizing and understanding different perspectives, including those of other AIs. It's also described as "sympathetic" since it can comprehend viewpoints other than its own.

Technology Readiness Level	Level of Development
Theoretical	Provisional
Proof of Pinciple	Aspirational
Deployment at scale	Undetermined
Biomimetic	Undetermined
Time Horizon	**8 years**

S4: Shared (Super) Intelligence

S4 has the ability to work together with humans, other agents and physical systems to solve complex problems and achieve goals. This stage is called "artificial super intelligence" or ASI. It is the kind of collective intelligence that emerges when sympathetic intelligence works together with people and other AI. We believe that this kind of intelligence will come from many agents working together, creating a web of shared knowledge that becomes wisdom.

Technology Readiness Level	Level of Development
Theoretical	Provisional
Proof of Pinciple	Aspirational
Deployment at scale	Undetermined
Biomimetic	Undetermined
Time Horizon	**16 years**

CONCLUSION

At VERSES, we're developing a new type of AI based on the nested ecosystems of intelligence found in nature that will lead to ASI. Within these ecosystems, intelligent agents, both human and synthetic, work together to solve complex problems. Active inference agents make predictions, take action, and interact with their environment through the Spatial Web, which enables them to perceive and understand the world like we do. Instead of seeing humans as separate from or inferior to AI and the Spatial Web, our approach ensures that we remain integral participants. We are leaving behind the dystopian, sci-fi stereotype nightmares, for a future, where artificial intelligence evolves and enhances everything it touches, where Intelligence is woven into the fabric of our daily lives in the same way that electricity was in the 20th century. In this case, acting as the key to upgrade our civilization, address our greatest challenges and rise to our highest potential. We believe that an approach to digital intelligence formed of ecosystems, based on the fundamental principles found in nature and evidenced in neuroscience is the path to achieve these goals.