Evolution of AI

A history of algorithmic decision making

Evolution of AI

A history of algorithmic decision making

Jaap Vink

Author: Jaap Vink

Coverdesign: Jaap Vink

ISBN: 9789403877891

Contents

"Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful."[1]

[1] Box, G. E. P.; Draper, N. R. (1987), Empirical Model-Building and Response Surfaces, John Wiley & Sons.

1 Uncertainty, and the human quest for mechanical objectivity

The history of human civilization is, in its most fundamental sense, a history of decision making under conditions of uncertainty. For millennia, the central challenge of the human experience has remained constant. How to navigate a complex, dynamic, and often hostile environment with incomplete information.

From the Pleistocene hunter gatherer scanning the horizon for the seasonal migration of herds to the modern executive analyzing real time consumer data, the goal is the same: survival and advantage. The contemporary era is often characterized as the age of big data and artificial intelligence. The cognitive and structural foundations of these modern phenomena were laid thousands of years ago. The fascination with data driven decision making, predictive analytics, algorithms, and artificial intelligence is not a sudden break in human cognition. It is the latest iteration of a continuous process that began in deep prehistory. Faced with the growing chaos and complexity of urbanization, ancient societies sought to transcend the limitations of individual biological memory, processing capacity, and subjectivity.

To understand the evolution of artificial intelligence, the analysis must look beyond the silicon chips and code of the last century. This history is better viewed through a framework known as the OODA loop, “Observe”, “Orient”, “Decide”, and “Act”. Developed by military strategist Colonel John Boyd, this framework describes the recurring cycle by which organisms and organizations interact with their environment. Throughout history, humans have developed tools to facilitate and expand the capabilities of this loop. Humans invented tally sticks to expand the capacity to observe and remember. They developed statistics and algorithms to expand the capacity to orient and find meaning in chaos. Today, they build autonomous systems to execute the decide and act phases with speeds that surpass biological limits.

Predictive analytics serves as an overarching term for all algorithms and approaches that use statistical techniques and machine learning to create smart, algorithmic decision making systems. It is the practice of connecting data to effective action by drawing reliable conclusions about current conditions and future events. Within this umbrella sits the modern definition of an AI system, which the OECD defines as a machine based system that, for explicit or implicit objectives, infers from the input it receives how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Traditional computer programs follow strict rules written by humans, typically in an "if X happens, do Y" format[1]. AI systems are different because they can learn from data. They can figure out their own rules to achieve a goal and adapt to new situations without being reprogrammed.

Traditional predictive systems relied heavily on human coded logic, whereas modern AI systems vary in their levels of autonomy and adaptiveness after deployment. They represent a significant leap in the human ability to externalize the OODA loop. They do not solve the fundamental problems of subjectivity and uncertainty. The quest for a purely objective "God's eye view" of reality has largely failed. Instead, human history is one of building intersubjectivity, shared realities constructed through collaboration, specialization, communication, standardization, and normalization. This journey spans from the bone tools of the Ice Age to the transformer architecture of the 21st century.

1.1 *The analytical framework*

Before examining the historical impact of algorithmic decision making, it is useful to describe the framework that underpins this analysis: the OODA loop. John Boyd's model is a complex, non linear description of how cognitive entities survive and evolve. It describes a process of continuous learning and adapting with multiple feedback loops. The loop consists of four interwoven types of activities: Observe, Orient, Decide, and Act.

The Observe phase is the intake of information. It involves the

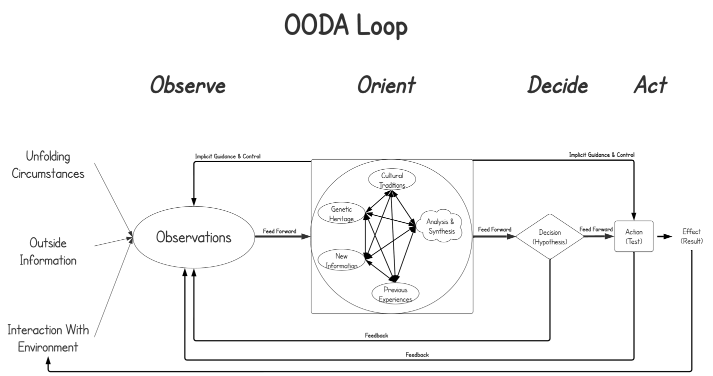

Figure 1 : John Boyd's OODA Loop

collection of observations from the environment through the senses or, in the case of modern systems, through sensors and data streams. It reflects unfolding circumstances, outside information, and feedback from previous actions. What humans decide to observe or what they filter out is determined in the Orient phase. Orientation is the most critical and complex phase. It is where observations become meaningful. It shapes the way humans interact with the environment, determining what they can observe and which observations they use, what actions they can take, how they decide, and how they act.

Boyd identified five key elements that synthesize to form human orientation: cultural traditions, genetic heritage or biology, previous experience, new information, and the process of analysis and synthesis. Modern cognitive psychology and neuroscience suggest that we must adjust this phase to include emotions and physical state.

Decision makers must also factor in externals that are not explicitly observed but do, unconsciously, influence decision making, like the weather, social environment, and personal situation. A decision maker who is fatigued, fearful, or biased will orient differently toward the same set of facts than one who is rested and calm. Orientation is the filter through which reality is perceived. It is where our mental models of the world are created, evaluated, and destroyed.[1]

Based on the orientation, a hypothesis or idea is formed in the Decide phase, or a course of action is selected. This is the transition from processing to intent. In the Act phase, the decision is executed. This action itself and its result impact the environment, changing the reality of the situation and generating new observations.

The OODA loop is not a one time process but a continuous cycle. The Act phase generates new observations, feedback, which confirm or deny the hypothesis or effectiveness of the action formed in the Orient phase. If the action produces the expected result, the mental model is reinforced. If it does not, a mismatch occurs. Boyd called the process of dealing with these mismatches deductive destruction and creative induction. Humans destroy old concepts and build new ones to match the observed reality.

The evolution of technology can be viewed as the progressive specialization and externalization of these phases. Writing externalized memory. Statistics externalized the analysis of patterns. Algorithms externalize the selection of choices. Robotics and automated systems externalized execution. Today, the digital landscape is designed to facilitate, support, and increasingly automate complete OODA loops.

1.2 Emotions and physical state

While Boyd’s original model was revolutionary, it did not fully incorporate findings from modern cognitive science and neuroscience that emerged after his death. Research in behavioral economics and psychology, such as the work of Kahneman and Tversky, demonstrates that human orientation is heavily influenced by heuristics and cognitive fallacies. Furthermore, the impact of emotions and the physical state of the body is now recognized as central to how humans process information and make choices.

Antonio Damasio’s somatic marker hypothesis provides a foundational insight into this process. Damasio discovered that emotions are not mere distractions to logical thought but are essential for effective decision making. Somatic markers are physical feelings in the body, such as a racing heart or a "gut feeling", that become associated with specific situations and their past outcomes. When a person faces a complex choice, the brain uses these markers to quickly filter options and guide behavior toward advantageous outcomes. Patients with damage to the ventromedial prefrontal cortex, the area where these markers are processed, often struggle to make even simple social or personal decisions despite having normal intelligence. They lack the emotional "compass" needed to navigate uncertainty.

Lisa Feldman Barrett further challenges the classical view of emotions with the theory of affective realism. Affective realism suggests that feelings influence the actual content of what humans see and hear. The brain is a prediction engine that constructs situated conceptualizations by combining sensory stimulation from the body and the environment with past experiences. From a neuroscientific perspective, perceptions are not objective recordings of the world. They are top down constructions biased by internal states. For instance, a person who is tired or hungry, a state of poor "body budgeting", may perceive an ambiguous social cue as a threat.

If the Orient phase is inherently subjective and tied to biological states, then any attempt to create a perfectly objective decision making system must confront the reality that humans are emotional decision makers whose behavior is naturally unpredictable. When humans build algorithms to automate orientation, they often aim for "mechanical objectivity" a form of detachment that avoids human interference. Yet, as we shall see, these algorithms are often just "fossilized human intent" that codify the very biases they seek to eliminate.

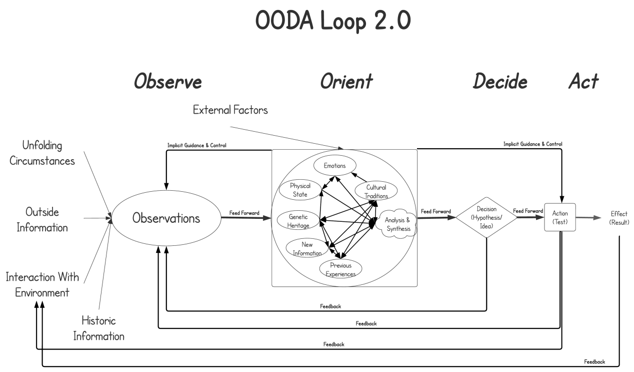

Figure 2 Updated OODA Loop

1.3 The mirage of mechanical objectivity

The quest to reduce subjectivity has led humanity toward the ideal of mechanical objectivity. Historians of science Lorraine Daston and Peter Galison[2] trace the evolution of this concept from the 18th century to the present. Early scientific practices followed a "truth to nature" model, where experts made informed decisions about which individual specimens were most "representative" of a species. However, by the mid 19th century, scientists began to distrust their own judgment. They turned to self registering instruments, cameras, and mechanical devices to produce data "untainted" by human bias.

Mechanical objectivity assumes that if a process is automated, it

[2] Daston, Lorraine, and Peter L. Galison, "Objectivity", (2021) Princeton University Press

becomes a neutral arbiter of truth. This belief is the foundation of the modern "trust in numbers". Theodore Porter[3] argues that societies turn to quantification and rigorous methods when they do not trust the people making the decisions. Numbers provide a "technology of distance", allowing decision makers to claim fairness and impartiality by following a priori rules. This drive for objectivity is a response to the need for managing large scale, complex societies where local trust is no longer sufficient.

Despite this pursuit, absolute objectivity remains structurally impossible. In physics, the Heisenberg Uncertainty Principle teaches that the act of observation alters the observed. In logic, Gödel's Incompleteness Theorems remind us that no system can prove all truths within itself. Every dataset curated and every algorithm written is an artifact of theory dependence. Humans collect only what they believe is worth measuring, based on the mental models of reality they hold. Measurement instruments themselves act as filters, capturing only a subset of reality. Believing in perfect objectivity is to deny the inherent subjectivity and bounded rationality of the human condition.

1.4 Externalization and automation

The history of humanity is marked by a drive to make the right decision at the right time. As societies grew more complex, the individual OODA loop became overwhelmed by the sheer volume of information and the speed of necessary interactions. This led to the progressive externalization and automation of OODA loop components.

Ancient civilizations like those in Mesopotamia, Egypt, and Rome were the first to implement "algorithmic governance". They standardized weights, measures, and legal codes to create shared realities. Scribes and priests acted as the first data scientists, maintaining census records and omen lists to tame the chaos of life.

---

[3] Porter, Theodore M., "Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. New Edition", (2020), Princeton University Press

These administrative systems functioned as externalized intelligence, allowing states to manage population statistics and fiscal capacity far beyond the memory of any single ruler.

The Industrial Revolution accelerated this process, placing data on "assembly lines". The 1890 US Census serves as a critical milestone, where punch cards and counting machines were introduced to prevent the collapse of the national census bureau under the weight of manual enumeration lists. This mechanization was not just about efficiency. It provided a veneer of objective, scientific legitimacy to the quantification of the population. These early data processing technologies also reinforced social hierarchies by embedding racial and demographic categories into the very structure of the cards.

In the 20th century, the shift from deterministic to probabilistic models marked a turning point in predictive analytics. Deterministic models follow fixed "if X then Y" rules, assuming the future is perfectly predictable. The real world, however, is fraught with randomness. Probabilistic models acknowledge this uncertainty, quantifying the "variance" or remaining doubt surrounding an expectation. This shift allowed for the development of modern algorithms, which infers propensities rather than absolute certainties.

Today, the automation of OODA loops is ubiquitous. In military contexts "Super OODA Loops" operate at speeds that massively outpace human cognition, providing a decisive advantage in high tempo environments. In business, algorithmic decision making has become the cornerstone of innovation. Organizations use algorithms to try to prevent psychological flaws like confirmation bias or the sunk cost fallacy by relying on data rather than intuition alone.

While automation speeds up the cycle, it introduces unique additional biases. Algorithms are not neutral. They are defined with goals in mind, such as efficiency or profit, which reflect human desires and greed. A rogue algorithm is often just a reflection of a "rogue psyche”, built on unexamined assumptions or data reflecting historical prejudices rather than current aspirations.

1.5 Limits of computation

A widespread misunderstanding of AI is the belief that computation can solve any problem if given enough data and processing power. Computer science and mathematics have proven that there are hard, immutable limits to what can be computed. These are not limits of technology, but limits of logic itself.

There exists a class of problems known as "intractable" or "NP complete". These problems are easy to describe and check, but practically impossible to solve perfectly because the work required grows exponentially. The Traveling Salesman Problem is the classic example. Finding the shortest route to visit a set of cities and return to the start seems like an easy problem. For 3 cities, it is easy. For more than 70 cities, the number of possible routes exceeds the number of atoms in the universe. Even if every atom were a supercomputer working for the age of the universe, the perfect answer could not be computed. This means that "perfect optimization" is mathematically impossible for many complex real world problems like airline scheduling, resource distribution, or protein folding.

Humans handle these problems using heuristics, intuition, and "good enough" guesses. If "Artificial General Intelligence" (AGI) is meant to solve any intellectual problem a human can, it must navigate these intractable problems. However, if it relies purely on computation, it hits a wall. This raises the question of whether human intuition is just a faster computer or a different process entirely.

Alan Turing proved the "Halting Problem". That it is impossible to write a program that can determine for any other program whether it will eventually stop or run forever. There is no "universal debugger". This suggests a "philosophical limit of reason". Logic is not a key that opens every door, and there are truths about the universe that computation cannot enter.

1.6 Deterministic certainty vs. probabilistic uncertainty

A key theme in this evolution is the shift from deterministic to probabilistic models. Deterministic models assume that the future can be perfectly predicted if the initial conditions and rules are known. Like a clockwork universe. They produce fixed outcomes: "if X then Y". The real world, however, is fraught with uncertainty arising from inherent randomness (quantum mechanics, weather), missing information, and the erratic nature of human behavior.

Probabilistic models, which underpin modern AI, acknowledge this uncertainty. They do not predict what will happen with 100% certainty. They predict what is likely to happen with a calculated degree of confidence. They quantify the "variance", the remaining uncertainty surrounding an expectation. As we explore the history, we will see humanity's struggle to move from the comfort of deterministic myths, like the infallibility of oracles, to the power of probabilistic science, like risk scoring.

This probabilistic nature of modern algorithmic decision making systems poses significant challenges and ethical concerns. Luciano Floridi's work on information ethics[4] provides a framework for understanding these issues, particularly in terms of agency, transparency, and ethical responsibility.

Some of the key issues with probabilistic models that Floridi identifies are:

Lack of transparency: models based on advanced machine learning often function as "black boxes", where the decision making process is obscured from users. This lack of transparency can lead to mistrust and skepticism. When users cannot comprehend how decisions are made, it becomes difficult for them to accept the outcomes.

[4] For example: "The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities", (2023), Oxford University Press and "Philosophy and Computing", (1999), Psychology Press

Uncertainty in decision making: the probabilistic nature of these models means that outcomes are not guaranteed but are rather based on propensities2. This can be particularly problematic in high stakes scenarios, such as criminal justice or healthcare, where decisions may significantly impact lives. Floridi emphasizes that the responsibility for decisions made by AI should not be neglected, as "the relationship between the degree and quality of agency that people enjoy and how much agency we delegate to autonomous systems is not zero sum". Users must be aware that these systems do not possess absolute certainty, which can lead to unintended consequences.

Ethical Considerations: The uncertainty associated with algorithmic decision making raises ethical questions regarding accountability and fairness. For instance, if an AI system denies a loan based on probabilistic assessments, it may inadvertently perpetuate biases present in the training data. Floridi argues that ethical frameworks must be developed to ensure that AI systems promote social good and do not harm vulnerable populations. The probabilistic nature complicates the establishment of such frameworks, as outcomes may vary widely depending on the context.

Impact on Human Agency: People may feel their agency diminished when they interact with systems that operate on probabilistic algorithms. The more these systems make decisions on behalf of individuals, the less control people perceive they have over their own lives. Floridi notes that thoughtfully developed AI can enhance human agency rather than replace it. However, if users are unsure about the reliability of AI decisions, they may begin to defer too much to technology, thereby undermining their own autonomy.

The implications of the probabilistic nature of algorithmic decision making are widespread. Firstly, there must be an increase in transparency and explainability. People are more likely to trust systems that can provide rational explanations for their recommendations, thereby fostering greater acceptance of algorithmic applications in various fields. Secondly, policymakers must develop regulatory frameworks that address the ethical concerns surrounding algorithmic decision making systems. This includes ensuring that algorithms are used responsibly, with attention to issues such as bias and accountability, which are exacerbated by probabilistic uncertainty. Finally, there is a pressing need for educating decision makers and users about the capabilities and limitations of algorithmic decision making. Understanding that algorithms operate on propensities can help individuals manage their expectations and engage more critically with these technologies.

1.7 Certainty vs. scope

Luciano Floridi has formalized another critical limit in algorithmic design: the tradeoff between certainty and scope.[5] This conjecture explains why building an AI that is both "general" and "reliable" is mathematically impossible.

Imagine a seesaw. On one side, you have epistemic certainty, how sure are we that the algorithmic is right? On the other hand, you have mapping scope. How much of the world can the algorithm handle?.

Epistemic certainty (C) is the probability that the AI system makes zero errors. A standard calculator has a certainty of 100%. If you type 2+2, it will never, ever tell you the answer is 5. It is perfectly reliable. Scope (S) measures the richness of the data the AI system deals with. A calculator has a very low scope. It only knows numbers. It cannot handle text. It cannot write a poem. Floridi’s conjecture is that you cannot maximize both at the same time. The formula is roughly: $1 C(M)*S(M) \geq k$, where k is a constant limit. This means as you increase Scope, Certainty must decrease.

Floridi’s conjecture suggests that as you increase the scope of an AI system to cover complex, unstructured domains, perfect certainty about the reliability of its results becomes unattainable. This explains the two main eras of AI history. In the 20th century, scientists tried to build "Symbolic AI" using strict logic rules. These

[5] Floridi, Luciano, (2025), “A Conjecture on a Fundamental Trade Off between Certainty and Scope in Symbolic and Generative AI”, Philosophy & Technology

systems had high certainty but tiny scope. They could play chess but crashed if asked about the weather. They were "brittle". They couldn't handle the messy real world. Today, with advances in machine learning, the focus has swung to high scope. These systems can talk about almost anything, but they necessarily relinquish the possibility of zero error performance. This is why LLMs "fabricate"[6] or produce factually false content. They are operating in a high scope regime where certainty is sacrificed for generality.

This trade off is a dagger in the heart of the "Artificial General Intelligence" (AGI) dream. The popular idea of AGI is a machine that is both general, has a high scope, and superintelligent, provides high certainty. Floridi's analysis suggests this is a contradiction. If you want a machine to handle the open world, it will make mistakes. It will be probabilistic, not logical. It will be fallible, just like a human. If you want a machine that never makes mistakes, it must be restricted to a controlled, limited environment, like a calculator or a chess bot.

1.8 Complexity vs. difficulty

If AI is bound by such hard mathematical limits, why does it seem so powerful in modern life? The answer lies in the concept of "enveloping" the environment. This requires a distinction between complexity and difficulty.

Complexity refers to the computational resources, steps, memory, and speed required to solve a problem. Playing Go or Chess is incredibly complex due to the billions of possible moves, but the rules are clear and the environment is fixed. Machines excel at complexity. Difficulty refers to skill, motor control, and dealing with the chaotic, messy, analogue world. Tying a shoelace or folding laundry is computationally simple but physically difficult. Robots are often clumsy and struggle with difficult tasks that a toddler can

---

[6] I prefer "fabrication" to "hallucination" because LLMs and other Generative AI models are assembly lines for content. "Hallucination" is too anthropomorphic.

perform easily.

The secret of algorithms' success is the translation of difficult tasks into complex ones through enveloping, wrapping the environment in a protective, machine readable layer.

Instead of building a robot that stands at a sink and scrubs dishes, a difficult task requiring hand eye coordination, humans built a box. They changed the environment. Dishes are racked in a specific grid, reducing chaos, and blasted with water in a contained process. The dishwasher is not "smart", but it solves the problem by enveloping it.

Another example is the self driving car. Truly autonomous driving on a 1960s road is too difficult because the robot cannot predict the chaos of human behavior. Success comes from enveloping the road, painting bright white lines, putting chips in traffic lights, and mapping every inch in 3D. The messy road is turned into a structured "video game level" the machine can navigate.

1.9 Creating shared realities

If mechanical objectivity is impossible, how do complex societies function and collaborate? They do so through the creation of intersubjectivity, the perceived commonality of thoughts, beliefs, or inner states with others. Intersubjectivity refers to a shared perception of reality that has evolved through social interaction and communication.[7]

Quantification and standardization are powerful tools for building intersubjectivity. The history of the metric system illustrates this process. Before the 1790s, European marketplaces were a "jungle" of units. Every town and trade had its own measures for length and weight, often based on the dimensions of a specific ruler's body. This inconsistency was more than a nuisance; it was a source of unfair trading practices and social frustration. The French Revolution provided the opportunity to replace this chaotic situation with a rational system based on "nature". By defining the meter as one ten millionth of the distance from the North Pole to the Equator, scientists created a standard that was "for all people, for all time".

This was not a discovery of objective truth, but the construction of a shared reality. Standardization allows for international trade, scientific collaboration, and large scale manufacturing because it creates a shared language of measurement. Similarly, bureaucratic systems like the census and legal codes provide a shared view of reality that enables communication and coordination across a complex society.

This shared reality is often biased. Standardized tests, developed to treat all students equally, often base their questions on standards pervasive in society, favoring those from the dominant cultural background at the expense of others. This creates a "magnifying glass effect”, where those who do not fit the "normal" profile are put under intense scrutiny. Intersubjectivity is thus a double edged sword: it enables the functioning of complex society, but it can also codify and naturalize social hierarchies, disguising them as objective facts.

1.10 The shadow of the algorithm

The drive to automate OODA loops through AI brings significant risks and ethical challenges. One of the primary dangers is "automation bias", the tendency to trust automated systems too much. Research shows that even experts often defer to an algorithmic, doing less independent thinking as a result. This leads to two types of errors. Omission errors occur when we fail to act because the AI system did not provide an alert. A commission error is the result of following a wrong AI recommendation even when contrary evidence exists. Recent failures in the Netherlands demonstrate the real world consequences of these biases.

In 2013, the Dutch Tax Authority introduced an algorithm to profile benefit claimants for fraud. The model was optimized to identify individuals similar to those already suspected of fraud, rather than identifying fraud itself. It developed a bias, focusing disproportionately on people with the very lowest incomes and those with foreign backgrounds. Officials trusted the algorithm's profiles without meaningful intervention or questioning. They interpreted probabilistic outcomes as deterministic, ruining the lives of thousands of families by flagging them as fraudsters without substantial audits. This case illustrates how a lack of transparency and a blind trust in "objective" algorithms can lead to grave systemic injustice.

In 2023, it was discovered that nearly 98% of students inspected for possible fraud by DUO, the Dutch agency for student finance, had a migration background. DUO used a simple rule based algorithm to flag students who might be wrongly claiming higher stipends for living away from home. While the system did not explicitly use ethnicity as a factor, it used "proxies", such as education type and distance from parents, that indirectly singled out minority students. Manual selection by staff further amplified this bias. Students with immigrant backgrounds had a three times greater chance of an unnecessary house visit than native Dutch students. The "magnifying glass effect" ensured that fraud was mainly found where the agency was looking, while ignoring potential fraud in other groups.

These examples highlight the "responsibility gap" in algorithmic decision making. When an algorithm driven decision causes harm, accountability is often blurry. Developers may blame the users for not overseeing the system properly, while users may say they were just following the computer. Policies requiring human oversight often fail because they treat it as a checkbox rather than a meaningful interaction. For oversight to be substantive, humans must be informed, capable, and empowered to override the algorithm’s results without penalty.

1.11 The “AI illusion”

As algorithms and AI systems continue to proliferate, it is vital to discern between the "illusion of intelligence" and the reality of predictive analytics. Anthropomorphic metaphors like "learning”, "seeing”, or "understanding" can be misleading.

At its core, modern AI is a computational process, not a reasoning mind. Large Language Models function as "stochastic parrots”, stitching together statistically likely words based on patterns in their training data. They lack semantics, the meaning behind the words. A machine does not know what a "moon" is; it has no sensory experience of light and no concept of a celestial body. Because it lacks a "ground truth”, it cannot distinguish between factual information and fabrication.

Algorithms are also structurally incapable of being "good" or "moral" in the way humans understand ethics. Moral judgment involves character and intent, "the agent", and the nature of the action itself, "the deed". An algorithm is purely consequentialist. It excels at calculating outcomes and optimizing for a "greatest good", chosen by its designer. It has no internal life. A "Care-bot" might mimic happy sounds when touched, but it does not feel love or compassion. In high stakes scenarios, an algorithm might compute that amputating a leg is the most "efficient" way to save a life, while a human would grapple with the horror and trauma of the deed itself Ethics requires qualitative engagement and "skin in the game”, making it non computable.

Finally, the increasing prevalence of algorithmically generated content risks creating a "recursive loop" or echo chamber. If the internet becomes flooded with generated text, future models will train on data produced by previous models. This could lead to a "homogenization" of content, where the unique nuances of human language are smoothed out into a statistically average "grey sludge". Errors and misinformation may become "snowed under”, compounded and repeated until original human perspectives are lost.

1.12 Meaningful “Human Algorithm Teaming”

The evolution of algorithmic decision making is not a trajectory toward a machine mind that will surpass human cognition in every dimension. It is the story of humanity's attempt to navigate a complex and uncertain world by externalizing its own cognitive processes.

The framework for the future is not algorithmic decision making replacing humans, but "Human Algorithm Teaming". In this relationship, the algorithm handles speed and complexity, while the human provides orientation, the meaning, context, and moral reasoning that the machine lacks. True wisdom lies not in the accumulation of more data, but in the awareness of the lens through which it is viewed.

To move forward we must abandon the myth of mechanical objectivity. We need to recognize that all data and models are theory laden and subjective. We need to ensure that every automated decision is open to review and reversal by an informed human. We need to design for contestability. We have to invest in algorithm literacy and critical thinking skills, educating people about the capabilities and limitations of algorithms and how to critically evaluate them, helping them to develop a calibrated trust. And finally, we need to embrace community in the loop oversight. We need to share the responsibility for algorithm governance among decision makers, designers, developers, users, and affected communities to ensure that pluralistic values are protected.

By standardizing and normalizing the world, humanity has built a complex society that allows for unprecedented collaboration. But we must always remember that the "shared reality" we have constructed is a choice. It is our responsibility to ensure that this reality is one that serves human dignity and justice, rather than one that automates prejudice under the guise of “neutral math”.

2 The birth of algorithmic decision making

We often think of algorithmic decision making, predictive analytics, and artificial intelligence as modern inventions. The mental and social foundations for these tools were actually laid thousands of years ago. To understand the evolution of AI, we must look past the computer chips and see the larger story of how humans have tried to move decision making out of their brains and into tools.

Throughout history, humans have invented technologies to make the OODA loop faster, more accurate, and more automatic. We created writing to help us observe and remember. We developed statistics to help us orient and find patterns in chaos. Today, we build algorithmic decision making systems and "AI Agents" to decide and act at speeds that no human can match. This chapter explores that journey from the first marks on a bone to the complex artificial neural networks of the 21st century.

Long before the invention of writing, the "urban revolution", or the rise of the state, humans were engaged in forms of data collection, storage, and analysis. In the Paleolithic era, the OODA loop was existential. A failure to observe the subtle signs of the changing seasons or to orient correctly to the migratory patterns of prey meant starvation. To mitigate this risk, prehistoric humans began the process of externalizing data, moving memory from the biological brain to physical artifacts.

The bureaucratic, divinatory, and administrative systems of ancient civilizations, specifically those of Mesopotamia, Egypt, Greece, and Rome, functioned as early forms of externalized intelligence. Faced with the growing entropy and complexity of urbanization, ancient societies sought to transcend the limitations of individual biological memory and subjectivity. They attempted to achieve the unachievable: objectivity, a "God's eye view" of reality.

The history of algorithmic decision making is often presented as a linear progression from Greek geometry to European algebra to

Silicon Valley. In contemporary imagination, the "algorithm" is a creature of silicon and electricity, born in the mid 20th century laboratories of Turing, von Neumann, and Shannon. We conceive of algorithmic decision making as a distinctly modern phenomenon, inextricably linked to the digital computer. When we strip the algorithm of its electronic hardware and view it in its purest form, as a finite sequence of well defined, implementable instructions, typically to solve a class of problems or to perform a computation, we find that its history is far older. It is deep, ancient, multicultural, and global.

The archaeological evidence reveals a global convergence where different civilizations, facing similar problems of complexity, independently or through interactions, discovered the fundamental laws of computation. In India, they built a machine of words to preserve the divine. In China, they built a machine of numbers to preserve the state. In the Andes, they built a machine of threads to preserve life. Claiming that all innovation came from western sources is misguided.

The ancient world's struggle was to close this loop effectively, to observe accurate data amidst the noise of the natural world, to orient through cultural and "scientific" models that filtered this data, to decide based on logic and calculation rather than whim, and to act to shape the future. The scribes, priests, and generals of antiquity were the first data scientists, constructing the initial algorithms designed to tame the chaos of the human condition.

2.1 Prehistory: first datasets and survival

The invention of externalized memory in prehistory represents the singular moment human cognition leaped from biological instinct to algorithm driven strategy. By carving notches to track lunar cycles or painting phenological[7] calendars, such as the dot and line sequences that track animal mating seasons, our ancestors created the world's first predictive models.

[7] Phenology is the study of plant and animal life cycles in relation to the seasons.

These were not merely artistic expressions but functional databases that externalized complex patterns of time and nature from the fragile "wetware" of the human brain onto durable "hardware". This shift was revolutionary because it decoupled information from the bottleneck of oral tradition. For the first time, critical survival data, when to hunt, where to migrate, or how to track resources, could be stored, verified, and shared asynchronously without the degradation inherent in "mouth to mouth" transmission.

This standardization of knowledge laid the cognitive architecture for algorithmic logic, “if pattern X, then outcome Y”, and predictive analytics, serving as the distant ancestral root of modern algorithms by allowing societies to make decisions based on recorded history rather than fleeting memory

The earliest evidence of data storage, the "hard drives" of prehistory, appears in the form of notched bones and tally sticks.[8] These artifacts represent a cognitive leap: the separation of information from the “knower”, allowing data to persist over time and being shared, creating the first glimmers of intersubjectivity.

The Ishango bone, discovered in the Democratic Republic of the Congo and dating to approximately 20,000 BCE, is a seminal artifact in the history of data processing. The bone features a sharp piece of quartz affixed to one end and three columns of ordered engravings. Some interpretations suggest it is a mathematical tool involving prime numbers or a simple tally stick, but the most compelling analysis for the historian of decision making is that it functioned as a lunar calendar. By tracking the phases of the moon, early humans could predict seasonal changes and animal migrations

---

[8] What follows is a common interpretation in archaeology. We do not have direct evidence. It could be that some of these artefacts are just the result of people ‘doodling avant la lettre’. This is also not exclusively related to homo sapiens. Neanderthal notched bones, such as the 51,000 year old engraved giant deer bone from Einhornhöhle and a 40,000 year old raven bone from Crimea, provide evidence of symbolic behavior and potential early counting systems.