Proof of Achievement of the First Artificial General Intelligence (AGI)

Daniel Olsher
Integral Mind (previously Carnegie Mellon, US Dept of Defense, Intelligence Community, DARPA, IARPA, State Dept, Singapore NUS Temasek Labs/DSO)
dan@intmind.com

Preprint Draft For Reference Purposes Only

Abstract

We have created the first-ever Artificial General Intelligence (AGI) and first superintelligence. Extensively proven for the US Government and in real-world application over many years, this paper provides the first detailed explanations of how and why the system works and conclusively proves that it does in fact deliver true AGI. After first deriving the requirements for real-world AGI from first principles, the paper sets forth the techniques required for AGI achievement and presents the first-ever definitive test for AGI - the Olsher Test. It engages key issues such as provable safety and responsibility and overcomes the work of Fjelland, Dreyfus, and other scholars who have previously argued that AGI can never be realized. It then further updates Newell and Simon's Physical Symbol System hypothesis and learning theory for the modern era. Finally, the paper explains a key corollary of the present work – that traditional approaches have been proven incapable of ever achieving AGI.

1. Introduction

We have created the first-ever Artificial General Intelligence (AGI) and first superintelligence. Extensively proven for the US Government and in real-world application over many years, in this paper we explain how and why the system works and prove that it does in fact deliver true AGI.

The paper begins by deriving the requirements for real-world AGI from first principles. It then sets forth the techniques we developed to solve AGI and presents the first-ever definitive test for it - the Olsher Test. It engages key issues including provable safety and responsibility. It then further shows how Fjelland's 2020 Nature Comms paper and the contributions of Dreyfus and others arguing that AGI is impossible are in fact entirely overcome by this work.

Finally, it updates Newell and Simon's Physical Symbol System hypothesis and learning theory for the modern era and proves that traditional approaches will never be able to reach AGI - even with unlimited investment.

2. How We Achieve AGI

In this section we explore AGI and how we achieve it.

2.1. AGI Core Requirements

To get to AGI we begin by examining the phenomenon at the core of the enterprise - *intelligence*.

Intelligence means succeeding at goals for which there are no rules in environments you have never seen before, cannot control, and that are constantly changing. (cf. Legg and Hutter 2007)

The key implementing mechanism for intelligence is *simulation*. Whenever an intelligent agent needs to think, plan, adapt to change, make a choice or decision, determine how to feel about an event, and/or accomplish any other traditionally 'cognitive' task, the best (and typically only) way to accomplish this is to ***accurately simulate the context within which we want to realize our goals***. We simulate the future created by each potential option so as to discover which ones create the best outcomes. Simulations not only teach us how we can achieve our goals but also what might go wrong, what risks we might face, and what we should do in response. They help us discover what our highest goals ultimately *should* be.

And thinking is best understood as a form of simulation; imagine deciding whether or not to marry a particular person. In our mind's eye, we'd simulate what our lives would be like with that person, how we think we would feel, and whether or not we think things would work out. We'd simulate every key aspect of life and ultimately make our decision based on where our simulations took us.

As we cover in much more depth below, the ability to *accurately simulate emotions* is essential to intelligence and AGI. But all emotions arise from *consequences*; it's impossible to know how to feel about something until we're able to imagine (that is, simulate) how it will affect us.

Understanding
Everything just discussed, and AGI, are impossible to achieve without *genuine understanding*. Real-world simulation, planning, adaptation, and success all require profound understanding of the world and how (and why) it operates. True intelligence *must actually work in the real world*, which is constantly changing, making the ability to deeply understand context (and others' mental states, which define core aspects of the real world) essential.

In the real world, anything *could* be relevant, so nothing can safely be ignored; *a true AGI must therefore be able to accurately understand <u>everything</u>*.

What is understanding?
According to the relevant literature, **understanding is fundamentally about grasping *why* things happen, in context, so you can use that information to support intelligent behavior** (cf. Baumberger 2014, 2019, Egler 2021, Gordon 2012, 2023, Khalifa 2013, Grimm 2021).

For this and other reasons, c*ausality* (understanding cause and effect) is therefore essential to reasoning, understanding, simulation, and explanation – everything at the core of AGI. According to Pritchard understanding requires "not merely being able to identify the cause" of an event, but also being able to "offer a sound explanatory story regarding how cause and effect are related. If one cannot offer such an explanatory story, then one doesn't count as having understanding, not even a limited understanding." (2014: 9)

Thinking and simulation both fundamentally involve asking 'What if?' in context. What if this happens? What if that doesn't? To answer such questions it is necessary to understand cause and effect. Accordingly, **any true AGI must necessarily be <u>fundamentally causal (and also not statistical) in nature.</u> Correlation is not causation.** 'Fundamentally causal' means that *all* implementing mechanisms must embody causality at their core – causality may not be simply 'bolted on'. As we explore in more depth below, *all* requirements for intelligent systems must be met for *any* to be properly met - any statistical aspects would prevent the rest of the system from functioning properly. A corollary of this is that no training data can be present.

Beyond this, any AGI is useless unless it can be trusted, which means that systems must be able to understand the world well enough to justify their conclusions, provide causal explanations, prove that they are correct, and help others understand why various paths are correct.

A corollary of the above is that any true AGI must be able to understand, simulate, and represent culture and the human mind; every real-world problem involves accurately understanding the thoughts, feelings, and motivations of others.

Context
Context is paramount. No real-world situation can ever be truly handled by rules; what is proper cannot be predefined (cf. Legg and Hutter 2007) because properness has to do with what makes sense in context and context is always changing. Plans are obsolete the moment they're created; any true AGI must be capable of leveraging its understanding of the world in order to continually adapt and adjust the outcome of planning processes in intelligent ways.

Indeed, everything in the real world takes place within, and cannot be separated from, some context. *The same action taken in different contexts will generate radically different effects,* which makes success and failure entirely dependent on context. This places context at the core of goal fulfillment, which in turn makes it critical to intelligence. Differences in context also lead to differences in explanations; people can only act with confidence when they understand *why* things are as they are, and in significant part it is a system's deep understanding of context that enables it to provide this understanding.

Contextual Adaptation
We just saw that true AGI demands constant adaptation in the face of change. But we never know what might change, or how, so we have to be ready for anything. ***Every aspect of an AGI must therefore support being changed in arbitrary ways*.**

Let’s consider some implications of this. Firstly, for change to be possible, we must never assume anything about the world within our system, as those assumptions would prevent change. Everything must always be fully flexible. Assumptions would also create strong, unpredictable biases that waxed and waned as the world changed, meaning that we could never *rely* on such a system.

When implementing change it’s essential to be able to change only that you want to change while leaving everything else alone. This means your system must be very *nuanced* – it must allow making very small changes.

Nuance is also important because you want the system to be able to understand the fine details of whatever is happening so it will be able to craft just the right response.

Accuracy

Reasoning, simulation, and contextual adaptation are key AGI capabilities. But if these aren’t *accurate*, they add no value. The terms in which the computer ‘sees’ the world must always match reality in all respects. This provides yet another reason why unlimited nuance is essential, as the smallest details of causality, context, and perception exert tremendous forces on outcomes. It is necessary to be able to understand *everything* – no system that ignores information can support true intelligence.

Exact, Proper Problem Solving

Proper problem solving is a key component of intelligence. A problem has not been properly solved unless we have a solution for the exact instance of the specific problem we face in the context in which we face it **and** we understand how the solution was derived and why it is correct. If the solution doesn’t match the problem, it can’t help us. If we don’t know how and why it’s correct, we won’t be able to trust it and won’t know when our solution has become obsolete and we need a new one.

Because traditional systems can only address specific tasks, it is often necessary to ignore nearly all aspects of problems in order to ‘fit’ them into what such systems can handle. Unfortunately, though, this typically means removing exactly those aspects that made those problems what they were in the first place. Traditional systems thereby end up addressing completely different problems than those being faced, often rendering their outputs entirely irrelevant along the way. In addition, because traditional systems are statistical, they can only provide *correlations related to the general class of problem* being attacked rather than *causal solutions tailored to the exact instance of the problem being faced in context*, which intelligence demands be the case. Because they are not causal, traditional systems cannot explain their solutions nor how they were derived. All of the hard work is left to humans, and it is never clear to what extent users can trust systems or for how long.

Each of the foregoing negative outcomes violates the requirements of intelligence and, thus, AGI. Systems with causality, understanding, and unlimited nuance, however, provide exact causal solutions and avoid all such issues.

Provable Safety, Responsibility

This AGI is provably safe and responsible. **It is never necessary to take anything on faith**, and **all desirable properties are present. The strongest possible proof proposition is offered.**

This system is entirely accountable to people, beneficiaries, society, policymakers, and technologists. Because it understands people and reality far more clearly than humans can, and understands psychology, it can often be far more empathetic than and can see and evaluate actions and consequences far better than people typically can. This also includes human potential – living one’s best life.

We offer safety and morality surpassing that of humans. The AGI is able to derive its own autonomous, independent moral and ethical judgements, without human input and in context, by simulating the effects of actions on real people. These judgments enable it to exercise self-control and limit itself in a provably safe manner.

These capabilities completely alter the AI safety debate, which to date has centered around weaknesses from which this system does not suffer and has ignored its unique strengths.

With respect to policy, this is the first AI capable of provably implementing and actualizing policy goals, including safety, fairness, transparency, and responsibility.

It is also the first AI offering immediate out-of-the-box compliance with all EU and US AI directives. It enables far stronger policy positions much more in line with what policymakers ultimately hope to achieve.

Lastly, as shown below, our specific implementation offers the **best possible 'safety floor' including optimal, provable safety, trustworthiness, accountability, provable nonbias, and complete transparency in all aspects.**

Safe Superintelligence
This system represents the first superintelligence as it is the first capable of thinking faster, better, more morally, and more optimally than human beings.

As the literature (e.g. Bostrom 2014) and our work for the US Government shows, **once any such system exists, everything is different from that point on.**

The computer makes much smarter and better decisions than humans would be able to do, so it is required that you trust it in order to succeed, as not doing so would necessary lead to worse outcomes.

Superintelligent systems shift all decisionmaking into regimes that move so much faster than humans that there's no time for people to double-check what computers are doing. The provable safety and provable self-control of this system are therefore essential.

This also makes the unique moral capacity of this AGI essential.

Superintelligence creates a race - those who don't have it won't be able to compete with those who do. Eventually everyone will need this - those who wait risk being left behind.

The endogenous morality and provable correctness and safety of this system enable us to obtain the real-world benefits of superintelligence in a demonstrably safe manner.

We believe it is imperative that AGI be deployed such that it always supports the strongest pro-human values. Doing so now places us on the right path while providing a bulwark against those who would seek improperly use this technology; **only good AGI is powerful enough to fight that which is bad**.

A core goal of this work is therefore to demonstrate how **superintelligence may be deployed in a pro-people manner**. **Path dependencies are critical – whoever deploys first sets the norm, and we are working to ensure that that norm is a wholly positive one.**

Because the power of AGI is unparalleled, so too must be its levels of accountability and trustworthiness be.

Only systems that can be actually proved to be safe and correct can be trusted enough to actually be deployed in practice. If a system *could possibly fail* in unpredictable ways, it's the same as if it never existed at all – you wouldn't be able to rely on it enough to give it the autonomy it needs to be useful. Autonomy is required, however, in order to compete with others' superintelligent systems. Moreover, in deploying unsafe systems we risk consuming all available resources trying (and ultimately failing) to control them.

In this way, **it is only when a system can be properly trusted (and provably so) that it becomes possible to generate the real-world benefits of this technology.** Hope is never enough.

Any AGI system must therefore be able to fully able to control itself, in context, exactly as a human would wish it to. It must know what limits people want it to have and be capable of keeping itself within them.

The AGI presented here is able to simulate the moral and practical effects of potential actions so is able to control itself in these ways. It knows what will happen if it does something and can determine the moral implications of those consequences, so is able to exercise fully trustable self-control.

It is therefore possible to prove that it will only ever act in pro-human ways.

This includes the ability to *teach and explain,* as explanation is essential. Humans cannot trust nor work together with a system unless they actually understand what is going on inside it. Systems must show people what is happening and why. People must be able to readily see and understand systems' inner workings so they can readily verify them when desired.

In order to generate this type of trust, it is necessary to prove that systems are correct at the top level (as we do). To support this, **our system provides full explainability and transparency across all aspects** – knowledge, reasoning, thinking, simulation, etc. It always explains why its responses are correct in context and can prove why any decision is a good or bad one. It teaches users what matters most in specific cases and can convey the limits of any recommendations that it makes.

Privacy - We Don't Want Your Data
This AGI has no use for, and its business model does not require, user data. **Contrary to the established wisdom, *data is not the new oil*.** Traditional AI needs data in order to train its models, but the instant system does not use models, doesn't train, and has no use for any such data. This enables further pro-human properties in that there are no inbuilt incentives to amass data and every interaction that does leverage user information can be entirely consensual, involve minimal disclosure, and can be revoked at any time. Users need only provide what they want to and no more. **There is no data cache to sell or protect.**

Provably Unbiased
In order to deliver all of the foregoing, any true AGI must be fundamentally and provably unbiased. It must be possible to show ahead of time that bias cannot and will not be present, as a biased system would not be useful and could not be safe nor trustworthy. Not only would it be impossible to predict where and how biases might manifest, but clear and unbiased understanding is required in order for systems to be able to think properly and conduct accurate simulations. Biases also necessarily prevent systems from properly delivering intelligent responses in certain situations, thus rendering them unable to meet core AGI criteria.

As we show below, the system presented here has identified and removed each potential source of bias in advance and its processes and procedures prevent bias from ever entering the system in the first place. At a minimum, because It is not statistical, does not use models, and does not employ training data, all biases resulting from these sources, including implicit human biases, are removed.

Independent Consciousness
In order to truly be free of bias, any genuine AGI must not rely on human cognition. Humans experience over 200 different categories of bias for which no remediation is possible. In addition, as we explore below, any level of human dependency renders AGI systems unusable in the real world and prevents them from being able to compete with other superintelligent systems in practice.

Reality is fundamentally too complex for humans to comprehend. In response, we evolved *consciousness*, which focuses limited cognitive resources on just those elements of situations that appear to be most relevant to survival while ignoring everything else. But this process creates profoundly false understandings of reality. What actually matters most in any given situation changes from moment to moment and always depends on context, but consciousness never quite catches up.

Consciousness causes us to see the world in arbitrary ways. It misunderstands and misconstrues much of reality and often suffers from confusion caused by viewing the world through stovepipes instead of as a whole. Assumptions build up, and while these are often incorrect, they are useful enough in everyday life, and are shared by enough people, that they tend to persist. But when making decisions, and when generating safely intelligent behavior, these assumptions damage every element of decisionmaking. They turn arbitrary accidents of history, process and personnel into key determinants of decision quality. Humans aren't aware of their biases and are unable to determine when their hidden assumptions and beliefs no longer apply. They often fall into 'tunnel vision'. As a consequence, traditional decisionmaking processes ignore far too much key information and are unable to adjust at the rate changing circumstances require. It becomes impossible to accurately see, predict, or understand, a consequence of which is that outcomes are dictated by chance, not choice.

Any genuine AGI must be able to see the world clearly and as it truly is, without dependence on humans and without bias, as only in this way will it be able to understand exactly what is happening (and why) and in so doing create order from chaos.

Biases sharply reduce the amount of information that can be accurately taken into account and introduce incorrect information, but this is entirely unacceptable if we wish to achieve the real-world success intelligence, and thus AGI, requires.

In addition, as noted above, **the mere existence of AGI leads to exponential leaps**. We all hold unconscious mental models that tell us what to expect from the world, **but our models in fact implicitly depend on the hidden assumptions that current levels of intelligence, speed, and human competency will continue to hold in future**. Superintelligence entirely shatters all such assumptions. Those who possess superintelligence are able to think, understand, decide, and act faster, better, and much more powerfully than everyone else, meaning that the human limitations implicit in current models no longer hold. The floor raises, and those who continue in old thinking are no longer able to compete. Once *anyone* has superintelligence, *all* must have it simply in order to maintain the status quo. Everything is therefore different from that point on.

The effects of superintelligence are especially confounding because they are strongly non-linear and take place across multiple stovepipes, meaning that no one person or group can easily see or understand them. Small changes in one part of the system exert major, unobvious, effects in apparently unrelated areas. People tend to hold on to their beliefs far past the time they have ceased to be useful, but this is no longer tenable in an AGI world. Anyone choosing to ignore new realities will be unable to win due to the immense advantages obtained by using better systems.

Indeed, only genuinely autonomous systems can compete on this playing field; human-on-the-loop and feedback-required systems are too slow, too biased, and lack essential properties.

Given the foregoing, the only possible conclusion is that **true AGI may not depend on human cognition, supervision, assistance, nor any human-in/on-the-loop mode of operation. As noted above, an AGI always works within its limits, and can always ask for permission, but it must have the autonomous capability to determine *how* those limits apply in any given circumstance as well as when to ask for permission.**

As a result, instead of depending directly or indirectly on humans, as all traditional systems do, **any true AGI must possess its own, fully and completely independent cognition** that does not rely on humans in any way. In addition, in order to meet the requirements of AGI, any such independent cognition must be implemented in such a way that it is possible to prove that critical properties hold of it. Independence can only be possible if **all system intelligence resides within the AGI itself *and never humans*. AGI may therefore never 'parasitize' human intelligence, as traditional systems must do.** As a consequence, any system that cannot think completely on its own, without drawing on content and/or decisions made by humans, cannot be said to have met the requirements of AGI.

Traditional systems parasitize human consciousness in many ways. Generative AI depends entirely on content created by human consciousness and possesses no actual intelligence of its own. All understanding continues to reside only in humans. Any system involving choices made by humans, including those that involve supervision (as LLMs and other systems commonly do) parasitizes human consciousness. Models depend on human choices with respect to their design and the decisions inherent in them, including what is important and what to leave out. They are top-down and cannot see any better than, or any different way than, humans. They fossilize human beliefs and biases and lack the intelligence and the understanding required to adapt. With respect to neural nets, these discover regularities in training sets *curated by humans* – that is, humans employ consciousness in order to decide what belongs within training sets and what does not.

An AGI's thoughts and perceptions must also be entirely independent. Combining these requirements, we see that ***any true AGI must have its own consciousness***. To achieve this, and to properly and accurately match the real world, true AGI must not rely on models nor any other type of structure intentionally created by humans. Creating top-down models implies choosing in advance how to look at things, but this creates profound bias, non-adaptivity, and rigidity, all of which prevent the achievement of AGI. In order to build understandings that exactly match the world and the problems being solved, and in order for systems to be able to perfectly adapt to these, **any true AGI must be able to independently build up its own unbiased picture of the world *without human guidance***. **Notably, this must be done from the bottom up**, as only in this way is it possible to avoid bias, enable unlimited nuance, and permit the world to be its own best model.

Independent Mind

This AGI has a mind because it does what minds do (cf. Searle (1980: 417) in Fjelland 2020). It understands all relevant causal information necessary in order to solve problems properly in view of user goals and broader context. Its purposes are those of the people it serves (to the extent such purposes are fundamentally compatible with the system's core moral compass). It leverages thought and simulation in order to find the best ways of achieving complex goals in context, and the causal nature of the system enables it to generate the steps necessary in order to put plans into practice.

Understanding the Human Mind

Any true AGI must be capable of accurately simulating the human mind. Much of what AGI systems are called on to handle depends in profound ways on specific details of how human minds operate, including meaning, the unconscious, activation, priming, associative memory, nuance, encyclopedic knowledge, local optimization (e.g. 'greedy' cognition), contextualized semantics, cross-domain reasoning, and other phenomena. While this has often been viewed as an impossible task, extensive published and unpublished work, including demonstrably proven work for the US Intelligence Community, has shown that it is entirely possible to simulate all of these phenomena via the mechanisms presented here (cf. e.g. Olsher 2013b, 2015, Olsher and Toh 2013, Olsher 2013c, 2013, 2014). This past work has demonstrated that simulations of the unconscious are essential to understanding and predicting human feelings and behavior and that, when properly taken into account in our system, vast predictions become possible. For all of these reasons, it is essential that any AGI be capable of representing, reasoning about, and simulating such phenomena.

Learning

'Machine learning' is in truth a misnomer, as traditional systems aren't genuinely *learning* – only collecting correlations. The relevant literature suggests that true learning should best be viewed as a process of increasing causal awareness such that we are able to act more effectively in the real world (cf. De Houwer, Barnes-Holmes, Moors 2013). Adapting this to AGI, we can define true learning as increasing our awareness of causal regularities in the world and encoding these in a manner that enables systems to generate perfectly context-adapted intelligent behavior at runtime.

Correlations are not powerful enough to support AGI. AGI demands deep causality and correlation is not causation. Correlations don't contain enough information to support intelligence, as all we know is that they held at some point for reasons that are hidden from us. We never know why they arose, what we can learn from them, when they might be applicable, or when they may stop applying in future. In this sense, correlations are unable to *teach* us anything, so we cannot *learn* from them.

It this connection is critical to note that, for perfect contextual adaptation to be possible, when encoding causal regularities nothing can be left out and no worldviews may be imposed as this would prevent systems from making different choices later on. We cannot change things based on what we think makes sense or encode predictions of the future, as perfect adaptation requires full freedom. Traditional systems inevitably fail on this point; the products of human thinking cannot be separated from the context from which they arise and thus inevitably embed all biases hidden therein.

Our processes and procedures enable us to build knowledge fully meeting the above criteria.

Language

Any true AGI must be capable of understanding and processing language as humans do – that is, in a genuinely meaning-based, not statistical, manner. Language is fundamentally about understanding, simulating, and building meaning, and genuine AGI must be capable of supporting these processes in full depth and with full contextual sensitivity. They must also be able to handle the human 'cognitive quirks' described above, as these greatly affect how language works and how it must be processed.

Generality And Strong AI

Generality in AGI is a phenomenon that emerges out of what a system is, what it does, how it is designed, and the properties and capabilities of its core components. As we will see, much like intelligence, we can only truly understand generality as *the confluence of multiple contributors, all of which must be present* in order for the phenomenon itself to exist.

Intelligence means being able to successfully solve problems we've never seen before in ways that make sense in the real world (which is always changing). To do this, a system must be capable of solving any potential problem at any time. If a system can't cope in any respect it will readily be defeated by others that can, thus violating the core of intelligence. The world won't limit itself to only those problems a system is able to solve, and it can't expect others to hold back. *There's no time for a system to* *catch up.* Every problem we face could potentially depend on the successful real-time solution of others (which could in turn span multiple domains). Additionally, as shown above, a system that can't understand, can't meaningfully adapt, is somehow inferior to human intelligence in any way, or isn't moral/transparent/provable enough to be trusted cannot possibly deliver true intelligence. Taking these together, we see that any system capable of genuine intelligence must be fundamentally capable of understanding anything, solving any problem, and simulating any phenomenon.

This is the first requirement of generality – a system's fundamental characteristics must be provably capable of handling all problems, contexts, and domains.

We include 'provably' above because, as we have seen, safety, real-world deployability, and usability all demand that a system prove in advance that it will work for all future cases and that all properties will continue to hold. Intelligence also demands provability.

As a further condition of generality, **it must be possible to prove that a system will be able to solve any new problem without the need for additional fundamental research**. We can see that this is true because the mere existence of any area requiring further research is enough to show that a system's fundamental characteristics are not currently powerful enough to work in all cases. Beyond this, any area requiring further research would by definition be one in which the system is not currently fully functional, and generality demands that no such areas exist. In practice, **no more than minimal reconfiguration and reasonable knowledge building may be required in order to shift an AGI system from one problem to another**.

True AGI must support all aspects of intelligence. Woodrow notes "the capacity to acquire capacity" as an essential component here (in Legg and Hutter 2007: 7). Because we cannot predict what sort of capacities we might be called upon to acquire in future, we again see that it is necessary for systems to be able to handle all potential problems, contexts, and domains.

With respect to generality, provable safety and other properties demand that no faith be required. The only way to prove generality such that we can be sure that it will hold for all future cases, regardless of application, is to ***prove that the necessary properties hold at the system level*. Proof by property is the strongest possible proof method and the only one strong enough to prove generality**. This is the only method capable of teaching us anything about problems we've never seen before, as properties of the whole cannot be proven from individual instances. Successful applications cannot tell us anything about whether or not the underlying principles of a system are powerful enough to deliver what AGI and intelligence require; thus, attempting to infer global properties from worked examples causes us to focus our energies in the wrong areas. The basic principles underlying a system must be shown to be powerful enough that generality has clearly been achieved.

Elegance and Property Floors

In addition to the foregoing, **only systems that meet a certain level of elegance can deliver generality**. **Inelegant aspects generate weak systems**; any aspect of a system that would make it inelegant would also make it incapable of delivering AGI.

As we show below, the system presented here demonstrates ***maximal elegance* in that each aspect of the system is ideal.** It therefore sets an 'elegance floor' that any putative alternative AGI would also need to meet (though a corollary of the foregoing is that no other approaches to AGI are likely to be possible). Elegance is therefore not just a 'nice to have' - something we might hope to find within the 'holy grail' of AI - but is in fact necessary in order for it to be achieved.

The system also offers ***all potentially desirable properties***, thereby establishing a second floor in this regard.

What is elegance? While the Olsher Test provides the definitive reference for this, it is useful to review some of the core elements here.

Fundamentally, what elegance would not admit an AGI may not depend on.

A core aspect of this is the complete avoidance of the need for any form of 'cheating', which we define here as the use of principles weaker than those required for genuine intelligence and AGI.

Cheating notably includes any use of statistics, therefore immediately excluding all traditional AI approaches and LLMs from ever being able to achieve AGI (we explore this in more depth below). This is true for multiple reasons. Firstly, true problem solving requires causality, but statistical systems are not causal and therefore cannot solve problems directly. Instead of attacking problems at their core, statistical systems are limited to processing ancillary data points (that is, items that, for reasons unknown to us, appear to correlate with key aspects of problems) without awareness of context. Thus, instead of actual solutions we receive decontextualized correlations to aspects of solutions which we cannot trust. And the use of ancillary elements in and of itself (instead of tackling problems directly) represents a form of cheating. Traditional systems attempt to find and apply useful correlations, but there are never any guarantees. We don't know what caused correlations to hold in the first place, so it's impossible to know which of them to rely on or for how long. All of the hard work of turning correlations into outcomes, and the entirety of the risks that come with such approaches, remain entirely with humans. Users are limited to making uninformed guesses and hoping that their statistical analyses might have some bearing on what they're trying to achieve. Traditional systems also deeply reproduce human biases. This untrustworthiness, use of human consciousness, uncertainty, lack of direct causal solutions, and lack of provable properties, at a minimum, all violate the core requirements of AGI.

And, critically, the ancillary elements that might correlate for certain problems don't work for others, meaning that each time we want to attack a new problem it is necessary to invest significant resources into finding elements and correlations that could potentially apply to that problem. This severely violates generality.

Finally, correlations also lack the information required to drive understanding, which is absolutely required for intelligence. In AGI, weakness *anywhere* is weakness *everywhere*. As a consequence, no genuine AGI may employ statistics in any fundamental respect.

Further indicators of cheating and lack of elegance include insufficient proof that systems are wholly unbiased, lack of explainability and/or transparency in any respect, dependence on human cognition, planning outputs that omit information about when/how they may become obsolete and what should be done in response, imperfect optimization in view of context and specific situations, planning that generates less than optimal success in the real world, anything less than real-time, context-aware intelligent adaptation, the inability to perform nuanced simulation, and/or the presence of any element not strong enough to meet the requirements of the Olsher Test set forth below.

Elegance must also include the complete avoidance of search, brute force, and/or any operation potentially requiring exponential time. AGI systems demand proof that properties will hold in all cases, but this becomes impossible in systems that employ search, as the specific details of what searches are able to find, whether or not those outcomes are optimal, whether local minima are involved, and so on, all depend on specific circumstances and therefore cannot be predicted in advance. This means that search, and therefore broader systems, may work in some cases but not in others, thereby violating necessary AGI requirements including at a minimum those relating to safety, trust, real-world usability/deployability, generality, and the requirements of intelligence. Search and brute force also prevent systems from providing the instant, time-predictable responses required in order to be able to act in real time and successfully compete with other systems.

Systems that employ search are also demonstrably inferior to those that don't require this, thus violating the elegance floor.

AGI systems are further prevented from employing *trial and error* as this prevents strong property proofs and violates requirements for predictable performance in line with the reasons given above. Trial and error implies the need to continue trials until thresholds are met, making systems entirely unpredictable as well as dangerously slow (that is, slower than competing systems that don't require this). As we show here, understanding, simulation, and causal reasoning capabilities, when properly designed, entirely remove the need for trial and error and/or search. Understanding immediately tells us what matters and what doesn't, and well-designed causal reasoning is capable of guiding us towards optimal solutions without ever requiring search.

Finally, and for the reasons given above, systems must be capable of achieving optimization (including causal optimization and optimal choice determination) via maximally-elegant means.

The system presented here meets all of the foregoing requirements.

As a last observation, we note that all systems employing traditional machine learning (ML) and/or traditional top-down models will be unable to meet AGI elegance requirements. ML components introduce reliance on statistics and cannot support strong properties (thus foreclosing AGI achievement) and are also unable to support true learning as defined above. Beyond this, top-down models parasitize human cognition, render independent cognition/consciousness infeasible, generate systems that are unable to adapt, and violate multiple other fundamental requirements.

Note Re Strong AI

The requirements set forth in this paper, and the Olsher Test below, provide detailed blueprints as to what constitutes ‘strength’ in AI contexts.

Strong AI, generality, and AGI are all best viewed as components of a larger whole; each of these concepts overlaps with, provides support for, and co-creates the others. None may exist independently.

Practical Know-How/Specialist Expertise

A final core component of this system consists of an extensive body of practical knowledge derived from real-world experience with customers and system deployments.

Proper design, implementation, validation, proving safety and correctness, overcoming obstacles, working with stakeholders, and anticipating and meaningfully responding to complex human factors issues that inevitably arise as part of AGI deployments all require unique specialist expertise. Because this AGI and traditional systems operate from entirely different foundations, traditional AI knowledge cannot offer any assistance in this regard.

2.2. Brief AGI Requirements Summary

The following sections explain how we solve AGI. As the solution is best understood in light of the above requirements, we briefly recapitulate them here. The following summary is included for reference purposes only and is not intended to be exhaustive in nature.

At a minimum, a genuine AGI must:

- understand everything,
- understand, demonstrate deep awareness of, and simulate context,
- adapt to changing contexts and environments and generate changes in plans, simulations, understandings, and/or other phenomena appropriate to those changes,
- explain causally what is giving rise to something, what phenomena are, why they matter, how you would achieve various changes with respect to them, the effects of changes on them, how they interact with other things, and how specific changes to the current context would manifest with respect to those systems and those phenomena in context,
- properly apply knowledge in context-aware ways,
- demonstrably be able to answer complex counterfactuals in context,
- demonstrably reason in a nuanced, contextualized causal manner,
- generate adaptive behavior capable of meeting goals in diverse environments,
- understand, simulate, and predict phenomena relating to the human mind in a provably correct manner,
- possess holistic causal awareness and accurate, nuanced knowledge sufficient to accurately causally simulate phenomena, systems, and the phenomena that affect them, in context,
- possess sufficient information, properly represented, to enable a system to successfully and properly re-adapt information in response to context,
- prove, by properties, and not by examples, that system will work for all future problems in all domains,
- show that system is designed such that new problems will not require fundamental research, regardless of domain,
- support deep full-semantics language processing,
- enable proper separation of the two stages of learning and fully support the processes required at each stage,
- have its own independent autonomous consciousness, cognition, and ‘mind’,
- demonstrate optimal elegance in all respects,
- deliver all perfect properties with respect to AGI,
- provably meet all necessary and desirable obligations to society,
- deliver provable morality,
- exercise provable self-control,
- deliver provable safety across all aspects,
- include all practical know-how necessary to achieve project success in real world,

- demonstrate the usability of system in the real world,
- be able to handle incorrect information and add new information without losing previously-proven properties,
- be provably unbiased,
- work ideally in all new, changing environments,
- never depend on humans, never need human assistance or on-the-loop modes of operations,
- be safe to deploy autonomously,
- be capable of competing with other instances of this AGI and all future superintelligences,
- meet all criteria for true generality,
- never 'cheat' in any respect,
- be perfectly transparent, perfectly explainable in all respects,
- deliver Strong AI,
- avoid premature adaptation across all aspects,
- show that all aspects of system are open to being changed in arbitrary ways (so as to enable adaptation),
- support unlimited nuance,
- be fundamentally causal in all respects,
- be capable of teaching humans and enabling them to see and understand the world in new ways,
- solve exact problems being faced, never require problems to be changed to fit into system,
- deliver proper learning as set forth herein, including no premature adaptation in the first stage, and
- be capable of delivering each of the intelligent processes and prerequisites described throughout this paper.

2.3. How We Achieve AGI

We see above that deep understanding, accurate causal thinking, deep simulation, and other capabilities are required in order to achieve true AGI. But how do we get there? We must begin by considering the question of knowledge representation (KR) – the terms and manner through which a system views the world - which, as it turns out, is the key to the entire enterprise. To understand, think, and simulate the real world in an accurate and unbiased manner we need the right KR; everything important about an intelligent system depends on this.

Davis, Shrobe, and Szolovits' 1993 paper can help us understand why this is the case. As they note, a KR is first and foremost a surrogate for the real world – a substitute for everything that exists. The first thing we need, and the foundation for everything else, is a representation capable of accurately representing the nuance and semantic and causal structure present in the real world. If we don't have this, we're finished before we begin. A KR is "an answer to the question ... In what terms should I think about the world?" (p. 17) It determines what a system can 'say', think, simulate, and, ultimately, do. It determines what is natural (and is thus more likely to be done and done well) and what is not. It determines which properties, and which biases, are present, what is provable, whether or not the system is truly general, and so on.

With respect to knowledge representation, intelligence, AGI functionality, and real-world usefulness demand at least the following (at a minimum, summary is non-exhaustive):

- causality as fundamental,
- unlimited nuance and 'grain' of any size with respect to information and causality,
- full support for information that remains unadapted at all times, thus enabling real-time contextual re-adaptation,
- universality – must be able to represent *any* information in any domain, including the human mind,
- cross-domain integration – it must be possible to reason seamlessly between disparate domains and apply common techniques regardless of where information comes from,
- fully represent holistic wholes,
- properly enable two-stage learning and knowledge application,
- support eliciting and representing tacit knowledge,
- enable provable building and testing without damaging properties elsewhere in system,
- never parasitize human intelligence in any aspect,
- support the proving of all potentially desirable properties of the KR and systems built on top of the KR,
- be fully transparent and explainable,
- support provable lack of bias,

- never ‘cheat’,
- support full elegance across all aspects of systems built on top of the KR,
- maximum inference – must provably support all potential inferences achievable from knowledge bases (otherwise there will be certain things the system will not be able to ‘think’),
- monotonic performance under knowledge addition – it must be possible to add new knowledge to the knowledge base without disrupting the current performance and/or provable properties of the system,
- noise resistance – it must be possible in some instances to accept some proportion of noise in the knowledge base without damaging system performance and/or provable properties,
- tractability – system operation must not include any operations that could become intractable,
- lack of search – a system must never depend on search and/or any other operation that may or may not succeed, be optimal, and/or potentially be intractable or not provably correct,
- represent all forms of information in wholly unbiased fashion, and
- fully support all elements of the Olsher test below.

2.3.1. Atoms Solve AGI

The foregoing tells us a great deal about the kind of KR we need in order to deliver genuine AGI Intelligence demands accurate, context-aware understanding, simulation, and reasoning. To meet AGI requirements, our KR must be capable of delivering all of the foregoing while being provably transparent, trustable, provably correct, and provably unbiased. How can we achieve this?

As it turns out, we can achieve all of the properties listed above and accomplish everything we wish to in the ways we want ***by breaking the world into small pieces*. *We call these* atoms*, and they are the key to AGI.***

Firstly, atoms bring us **universal KR with unlimited nuance.** To imagine how this works, consider the art practice of ‘pointillism’ – building pictures out of dots. Any picture can be made out of dots; if you need more detail, simply add more dots. Size is critical here – if your dots are large, you’ll only be able to see ‘blobs’. It’s impossible to build a proper picture out of large blobs. **But with tiny dots you can perfectly match anything**. You can build any picture you want. And as the picture in front of you changes, small dots enable you to adjust only that which has actually changed.

Atomic representation is universal (able to encode any information) because any information in any domain can be decomposed into atoms no matter its nature or inherent complexity. Because the size (amount of information) stored within atoms is flexible, we can readily use them to represent information at any level of abstraction.

Because each atom only carries a very small amount of information, atoms are able to perfectly encode unlimited *nuance*, which enables atomic systems to exactly meet reality as it is and on its own terms. **Atoms thus provide a digital interface to the analog world.**

Because atoms are small, they are flexible, enabling full, exact adaptation to context and change. Accurate understanding, simulation, reasoning, and contextual adaptation demand that the terms in which the computer thinks accurately match reality in all respects. There can be no instance in which the world needs to be simplified or otherwise changed in order to fit into the system and nothing can be left out. it is necessary to be able to simulate everything – otherwise, the system cannot meet the bar of true intelligence.

Unlimited nuance also means that systems are able to exactly match problems on their own terms, ensuring they can be truly general and can properly address all problems without ever needing to ‘cheat’.

To achieve this a deep, nuanced understanding of the current environment, the motivations of others, and all other relevant factors is required. The smallest details of causality, context, and perception exert tremendous forces on outcomes. Ignoring any level of detail would inject fatally unrecoverable bias into the system. Context is always subtle and implicit; if a system can’t see all the fine details, something important will inevitably be missed, leading to uncontrollable bias, lack of safety, and systems unusable in the real world.

An insufficiently sensitive system would not be worthy of the mantle of strong generality. The ability to properly adapt to small changes in the environment is at the heart of intelligence. The small details make all the difference.

In addition, if we can represent anything, we can understand everything. This is a key AGI enabler.

Causality
Deep causality is at the core of AGI.

A key insight here is that we can apply the same thinking with respect to causality that we did above for information - if we set up our atoms to **store not just small amounts of meaning but also small amounts of *causality*, we gain all of the benefits described above (nuance, universality, provable correctness, etc.) within the causal realm**. If we further add a *direction* to these small bits of causality, we can represent ***cause* and *effect***.

Storing information in this way enables us to readily follow chains of cause and effect, meaning that **search is never required.** Full performance becomes possible with extremely low computational requirements regardless of the size of the knowledge base.

This deep support for causality enables the AGI to generate explanatory narratives and 'reason to the best explanation' (cf. Mueller 2006), both of which are essential to problem solving and simulation.

Context Independence and Contextual Re-Adaptation Support
As we discuss in depth above, AGI demands context sensitivity and the ability to adapt, which is only possible if all information that systems rely on remains independent of any specific contexts at all times. Because atoms are so small and transparent, they completely avoid the 'smuggling in' of any hidden context. Extraneous information is immediately obvious and can readily be removed. **Atoms therefore always remain entirely independent of context, enabling the runtime re-adaptation necessary for intelligence.** Because the representation doesn't lock in any particular view, the system is always able to adapt. The small size and direct accessibility of atoms makes it obvious when improper information is present and the fact that atoms must be obviously true to be correct (see below) ensures that nothing improper passes the gate. All elements of the system (knowledge representation, system design, algorithms, and so on) are therefore able to operate in a manner that does not prejudice adaptation in any way (thereby allowing full freedom at runtime).

Full Information Availability
By breaking it up and exposing it to the world, *atoms make all information fully available for use by systems* as opposed to hiding this within opaque symbols or otherwise impenetrable correlations. Systems are always able to access the detailed information located 'inside' complex wholes. We term this availability 'semantic surface area' (cf. Olsher 2014, 2013b). In traditional systems all knowledge remains in the world and is not in the system, so meaning is not accessible. But atoms bring information out of the world and into the knowledge base, making it fully available for use. This enables the system to *genuinely understand* the world as opposed to merely manipulating meaningless symbols or tokens, thus facilitating intelligent processing.

Contextual adaptation requires access to the high level of semantic surface area provided by atoms in order to do its work. Intelligent behavior is generally performed in service of specific ends, which requires converting the results of thinking, reasoning, and simulation into answers to questions and/or plans to be followed (and understanding of the conditions under which these would change). To do this, it is necessary to have access to the full range of information and nuanced internal details that atoms provide. Complete avoidance of premature adaptation and the system's fully independent consciousness allow this process to take place with full freedom – that is, fully unfettered by bias, pre-conceived notions, and so on.

All Optimal Properties
As noted above, the system presented here offers an optimal set of properties (all desired and potentially desirable properties) including provable correctness and full transparency and explainability across all aspects.

Fully Explainable
Fully Transparent
Atoms are completely understandable and inspectable by humans. All operations employing them are fully transparent and explainable, enabling stakeholders to readily observe the AGI's operations and know they are correct. It is straightforward to prove both semantic and causal correctness, and it can be readily proven that future operations can and will only be correct (see below).

Provable Lack of Bias
As we show in depth below, the optimal properties provided by this system include **provable lack of bias**.

All avenues through which bias could possibly enter have already been identified and entirely closed off. This remains the case at every lifecycle stage, from design to knowledge building to deployment and validation.

Every aspect of the system is designed so as to remove any potential source of bias from ever entering in the first place, and any putative biases would be immediately detected during review of knowledge and/or system operations.

Specifically, there are only two ways atoms can be 'wrong' – *incorrect atoms* and *lack of causal closure* (that is, the complete absence of some highly relevant causal element). Critically, ***our tools and procedures provably address all such concerns***. As we discuss in depth below, while it is in theory possible for 'unknown unknowns' to cause bias, in practice the system is so fundamentally robust at all levels that this also never occurs.

While the typical case is that all atoms will be proven correct, there are some scenarios (including adaptations of traditional knowledge bases) where it is expected that significant numbers of atoms will be incorrect. We fully support such scenarios (see Olsher (2014) for in-depth discussion). Atoms encode incredibly rich semantic, causal, and systemic structures which systems can readily leverage in order to filter out bad atoms and maintain the vast majority of interesting properties despite incorrect knowledge. The real world is an incredibly messy place – enabling useful decisionmaking in the face of overwhelming noise and complexity is a core competency of this AGI.

Provable Correctness
Atoms' small size enables, and provable nonbias demands, use of an 'obvious total agreement' verification principle in which all atoms must be *obviously correct* or they are deemed incorrect (see discussion below re human involvement here). If any controversy is present, the atoms at issue must continue to be broken down until their correctness becomes obvious. We have developed fully repeatable processes for achieving such agreement within short periods of time and with low effort on the part of those involved.

Minimum semantic entropy (storing as little information as possible within atoms) enables the system to bring waste entropy (incorrect information 'smuggled in' with other knowledge) to zero.

Because atoms are completely transparent, any putative residual biases would be immediately apparent (and it would readily be apparent how to resolve them). While Inspection alone is always enough to verify correctness and that all properties hold, observers can also readily conduct imaginary 'mind experiments' in order to properly convince themselves that all atoms are correct.

All relevant aspects of atoms can be proven correct, including *depth* (all potentially relevant causes are captured and there is causal closure – also see 'unknown unknowns' below), *precision* (all relevant nuances are present and accurately captured), *correct correspondence* (knowledge is obviously true in that it matches reality and is not biased in any way) and *breadth* (sufficient knowledge is present from all relevant domains in order to address all problems being solved by that particular instance of the AGI).

With respect to depth and causal closure, the required standard is simple awareness of the existence of all even potentially relevant causal influences (we do not need to know how, why, or when any influence might be relevant, just that it exists). The processes and procedures built into the system enable this criterion to readily be met and to prove with all necessary certainty that it has indeed been met.

Our methods surface all even potentially relevant causal aspects and readily enable us to properly demonstrate that this has in fact occurred. As with the above, while this may sound difficult to achieve, our protocols demonstrably achieve it with minimum effort (as has been repeatedly shown in the past). As always, and as we discuss below, there is and can be no human judgment here - we simply ensure everything is present and the system does the rest.

Human choices are never part of the system. As is always the case, human choices, beliefs, and opinions are always and only viewed as bias.

Taken together, the foregoing covers all possible ways atoms (and as we will see, by extension, systems) could be wrong, enabling provable correctness.

New Knowledge Addition Remains Safe
Unlike with traditional knowledge bases, all proofs that held true of previous knowledge continue to hold when new knowledge is added – **adding new knowledge does not damage existing capabilities.** We term this 'monotony under knowledge addition'.

Provably Correct Outputs
It is always possible to prove atoms correct. Our system is not probabilistic in any respect and the correctness of reasoning and simulation depend only on the atoms in the knowledge base. Once atoms are proven correct it is straightforward to show that the system's thinking, reasoning, simulations, and explanations, as well its future functioning, are and will be correct. Barring an incorrect implementation or physical system failure, the system can be relied upon to act sensibly and morally as expected.

This system includes extensively tested, US DoD-compliant verification protocols that readily enable all of the foregoing checks to be readily achieved in real-world conditions.

Provable Safety, Morality
As we show below, atomic systems offer provable safety and morality, including provably correct self-control and moral understanding/simulation. It is possible to show that all decisions will be sensible across all future contexts and that proper empathy and limits will always be applied. Such systems are able to determine when and why they should ask for permission and can convey this to humans. Outputs and operations are fully inspectable, explainable, and provable in all respects.

How Atoms Are Created
Atoms may be generated via automated processes, from the processing of crowdsourced information (as they are in COGBASE), or via human-involved protocols, as they are in the typical case.

It is important to ask, however, why and how knowledge-building modes that involve humans should not be considered as 'parasitizing human consciousness' (which, as we have shown above, no AGI may ever do). How can such human involvement not introduce bias? Are we somehow reintroducing rules in another form?

Our protocols explicitly prevent and rule out any introduction or parasitization of human consciousness. Humans are allowed to perform only two tasks with respect to atoms: generate them (subject to their being proved correct) and offer a binary (yes or no) opinion on whether or not individual atoms are obviously correct. If not obviously correct, atoms must be refashioned until this is the case. Nothing else is allowed - *any other human involvement is strictly forbidden*. As is always the case, human choices and cognition are forbidden from affecting the system and choices, beliefs, and opinions are viewed as entirely invalid and biased. There is no 'room' in atoms for humans to be able to 'smuggle in' any extraneous information.

Thus, when constructing atoms, humans ***are never allowed to make any choices, or use discretion, of any kind***. We cannot try to tell the computer what to do and cannot impose our experiences, worldviews, thoughts, or beliefs in any way. We may not act based on what makes sense now or we think will make sense in future. ***Only the computer is allowed to make any choices – and this, at runtime and in an entirely autonomous manner*.** An atom must be so simple that there is no 'room' for any belief or discretion. **We cannot choose whether or not to add or leave something out – if something *could* go in, it *must* go in**, even if we don't think doing so makes any sense.

During atom building, we are extracting raw information, and are verifying that information, but are never structuring nor giving any guidance on how to use that information, all of which remains entirely up to the computer. No rules can therefore be said to be involved.

In addition, humans are expressly forbidden from telling the computer to look at the world in any particular way, so there are no models involved. The computer accesses raw reality directly via atoms and builds its own view of the world, in context, entirely without outside assistance or interference.

'Unknown Unknowns'
While it is theoretically possible for there to be undiscoverable significant causes ('unknown unknowns'), as has been borne out by overwhelming experience, this is never an issue in practice. We consistently find that our protocols discover all potentially

relevant causal influences very early on. Our processes are capable of surfacing hidden causes, but this has never been necessary in practice. We find that it is always clear when causal closure has been achieved and we have never encountered any case wherein the scope of even potential causal influences could not be readily identified. If there is any concern, however, the solution is very simple – simply continue to add more atoms. Further knowledge is always welcome – if the system doesn't need it, it simply won't use it. Adding new knowledge does not disturb existing knowledge.

Systems and Systems Thinking

In the real world, everything affects everything else, making systemic thinking a precondition for engagement with reality and therefore an AGI prerequisite. Indeed, **the only proper way to understand any subject matter is as interlocking systems going far beyond mere propositions – that is, *holistically***. Genuine AGI can only function on top of fundamentally systemic substrates.

A key benefit of atoms is that they spontaneously self-interconnect such that systems automatically emerge without further effort.

Atoms thus give us systems thinking 'for free'.

Grimm (as cited in Gordon 2023) notes that understanding involves genuinely useful awareness of "system[s] or structure[s] [that have] parts or elements that depend upon one another in various ways." This can include "field[s] (e.g., genetics), topic[s] (e.g., justice) or [i.e. social] system[s] (e.g., a historical inheritance system) ... narratives (e.g., history works), ... [and] non-propositional representations" (Baumberger 2019)

Atoms naturally connect to another in ways that implicitly bring forth such structures; simple traversals are enough to reveal the systemic content underneath. Systems emerge from the knowledge itself and implicitly guide operations, enabling fully-nuanced systemic understanding and processing.

The Human Mind

Atoms are especially-well suited to represent, reason about, and work with the human mind, including emotions, psychology, semantics, and culture (see e.g. Olsher 2015, 2013, 2013b, 2014). This type of information has traditionally been viewed as something so ethereal and ill-defined that it could not possibly be represented or processed within a computer. But atoms do provide such capabilities in provably optimal ways.

Atoms make the content, semantic structure, and functioning of the unconscious amenable to exploration and processing. The functioning of human memory and related semantic processes fall out as epiphenomena of the emergent structures atoms create within knowledge bases. As noted earlier, AGI requires the ability to simulate various 'quirks' of the human mind, including the unconscious, activation, priming, associative memory, nuance, encyclopedic knowledge, local optimization, contextualized semantics, cross-domain reasoning, and others (see previous cites). All AGI-necessary phenomena can readily be explained, represented, and simulated via the mechanisms presented here.

In addition, atoms have been shown capable of readily simulating and predicting human thinking (conscious and unconscious), emotions, psychology, and behavior. There is a clear and natural fit between the representational capabilities of atoms and what we know about the human mind, and we have successfully simulated these aspects across a very wide range of situations.

Interestingly, given the deep extent to which this system's simulations have been extensively validated as matching the real world, there is in fact likely some type of homomorphism that exists between the way the brain operates and our system. Evolutionarily, there may not be any other mechanism capable of generating intelligence. This is likely to represent an especially fertile area for future research in brain science and psychology.

Provable Morality and Sociality

Understanding the human universe as it does enables this AGI to 'put itself in others' shoes' and deeply understand what the world looks like from their perspective, often far better than humans can. This enables it to deliver the provable safety and morality necessary for real-world deployment.

Past work (see e.g. Olsher 2015, 2013, 2013b, Olsher and Toh 2013, Olsher 2014, and US Government work) demonstrates the ability of the system to feel and think as real people do and apply this knowledge in order to solve problems, including but not limited to accurately predicting actions and behaviors.

The system is able to fully simulate morality in all respects, determine the most moral courses of action in context, communicate these to others, use them to drive planning and action, and exercise self-control in view of all moral, ethical, policy, and other relevant requirements. It understands what matters most to people. This means that it is able to be far more cognizant of the effects of its actions on real people, and on society, than people generally are able to be. This completely changes the calculus with respect to social impact and moral AI.

The AGI performs all moral evaluations entirely on its own. When violations are detected, it is able to explain the issues that have arisen and their origins. In such cases it can suggest and explain potential solutions (or show why there aren't any). The system is powerful and helpful enough, and understands enough, to actually ***teach*** **people, meaning it is able to provide a far better understanding of the moral landscape underpinning complex issues** than humans typically have. And **its deep understanding of how humans receive and process information enables it to present new information and new understandings in ways that are far better able to transcend cognitive biases** than traditional discussion-based methods. In this way, it is able to **produce far better outcomes**.

The system further **helps people exercise good judgment** by simulating the effects of various courses of action via its provably correct substrate and ability to compute real-world consequences, including but not limited to morality.

Universal Fusion

Atoms are able to represent information from any domain within the same universal format. They can also fuse information from highly diverse systems and databases. By collecting all available information into a single all-inclusive microcosm, they enable the AGI to holistically simulate the effects of diverse influences upon one another. In the real world, problems always involve information from multiple stovepipes, which makes them difficult for humans to solve. The instant system, however, sees and understands information from all stovepipes as an integrated whole, enabling it, and users, to operate with a proper view of the entire picture. Given that an integrated understanding of the holistic whole is a prerequisite for generating intelligent solutions, this fusion capability is essential to AGI realization.

Integration with Statistical and Other Heterogeneous Systems

Our system is purely causal, never statistical. In some cases, however, users may wish to integrate it with others that are probabilistic in nature and/or that do not carry the same properties as ours does, such as image recognition tools. In such cases atoms are flexible enough to enable full semantic and causal integration between heterogeneous systems. This can be especially useful within legacy environments. Such integrations are able to 'sandbox' the influence of probabilistic inputs so as to ensure that the overall system continues to maintain all necessary and desirable properties while properly taking all inputs into account.

Sub-Simulations

Certain systems, especially physical systems, would benefit from the integration of physical sub-simulations and scale models (cf. Waskan 2003) into the broader AGI. Atoms' fusion and universal representation capabilities enable these sub-simulations and scale models to be readily integrated with one another in a fully transparent, provably correct manner. During simulation, the AGI is readily able to determine that reference to one or more specific sub-simulations would be appropriate, generate proper inputs for these, translate output semantics, and fold translated outputs back into the flow of simulation in view of context.

Broadly speaking, generality demands that AGI systems be capable of implementing sub-simulations of any type.

Atoms and Learning

As noted above, genuine learning involves two phases (discovering/encoding of causal regularities and runtime adaptation based on these) separated in time. Proper adaptation and/or application will only be possible if the first phase is encoded in an entirely context-independent manner. Atoms readily provide the required encoding mechanism, enabling the results of the first stage of learning to properly remain dormant until runtime when they can be usefully applied in context. Atoms also enable regularities to be encoded with full fidelity and with all nuances intact.

Discovering What Humans Cannot

As noted above, atoms and this AGI are 'model-free', meaning that they are not dependent on humans in order to determine how to think about the world. Traditional ontology- and model-based systems rely on human assumptions and worldviews, resulting in biases that are impossible to detect and remove. They limit functioning to exactly and only those situations that designers were able to anticipate early on, making it impossible to adapt to new realities and emergent situations (known as 'brittleness' in the GOFAI literature).

In complex environments, humans employ heuristics and oversimplifications in order to reduce cognitive load. While these appear to help in the moment, they tend to persist long past the point at which they are no longer useful, causing decisionmaking failures. Because our system explicitly rejects models and heuristics, it does not suffer from and can help humans avoid false beliefs and tunnel vision. **Simulations enable the system to teach humans when conditions have changed and new understandings are required**. The system actively rejects the use of human beliefs and/or opinions in any way.

Because they are so flexible and decontextualized, atoms allow computers to reason independently without reference to any model. Because they embed no 'right way' of viewing the world, the systems they underpin are able to dynamically adapt to changing contexts without human intervention. Atomic simulations help discover useful options, third-order effects, and courses of action that, due to cognitive biases, information overload, and other factors, would most likely have been ignored by humans. The system can infer the benefits of each option in view of they affect everyone involved, enabling the system to find solutions that are optimal for all parties. This understanding of goals and objectives, coupled the ability to project the effects of actions into the future, helps the system suggest appropriate courses of action. It can start with the desired ends in mind and work backwards in order to determine how best to achieve user goals in context. ***This frees users from having to know (or guess) how to solve problems.***

Properly Solving Problems

Atoms enable users to solve the exact problems they face, properly, causally, and on terms that make sense for those problems. They represent problems at the exact level of nuance and complexity required to fully understand and accurately simulate and solve them. Unlike traditional systems, it is not necessary to simplify nor change problems in order to fit them into what the system can do.

Atoms Enable Full-Semantics NLP

Atoms enable all aspects of full-semantics Natural Language Processing (NLP), particularly through the lens of the cognitive and functional linguistics paradigms. Semantics, syntax, frames, and constructions are all readily represented and simulated by atoms. This AGI is uniquely capable of providing the necessary nuanced semantic, understanding, syntactic/semantic fusion, and cross-domain fusion capabilities necessary to implement these paradigms.

True comprehension (that is, involving true understanding and not merely the manipulation of otherwise meaningless symbols) requires profound capabilities. Meaning is never merely embedded within language - rather, based on common context and shared knowledge, comprehenders are guided by language in order to construct particular meanings. This process is impossible without the ability to understand context and knowledge, apply these de novo in each new situation, and reason and simulate based on all of the above.

Previous work has extensively explored the definition, implementation, and application of these ideas.

It is helpful to view human languages as interlocking systems of form and meaning; atoms enable these to be processed simultaneously. Olsher (2012b, 2012) demonstrate how atoms can bind syntax and semantics in support of fluent processing and understanding. Olsher (2012) and other past work have shown the value of the Radical Construction Grammar (Croft 2001) framework with respect to syntax and these cites show how fused semantic/construction-based parsing can be achieved within this paradigm.

Olsher (2014, 2013b) address key technology for semantic processing, Olsher (2012, 2012b, 2014) demonstrate the use of the system to understand and extract the meaning of syntax, Olsher (2012b) dynamically constructs meaning from language in a semantically-driven manner, and Olsher and Toh (2013) and Olsher (2013, 2013c, 2014, 2014b and others) compute cultural and other implications and discuss the intersections of the foregoing with semantics, cognitive linguistics, and other related fields.

Updates To Newell & Simon Physical Symbol System Hypothesis

A key implication of our work is that Newell and Simon's Physical Symbol System Hypothesis (PSSH) is, if not wrong, at least fundamentally incomplete.

The PSSH states that “[a] physical symbol system has the necessary and sufficient means for general intelligent action." (Newell and Simon 1976) But this is not true, as traditional symbols are not sufficient to achieve AGI. To save the hypothesis, it would be necessary at a minimum to redefine the concept of ‘symbol’ towards that of atoms and to import all of the necessary aspects noted in this paper, including contextual adaptation/lack of pre-runtime adaptation, nuance, causality, provable correctness and lack of bias, safety, and so on, as we have shown all of these to be obligatory prerequisites for AGI.

As Olsher (2014) points out, traditional symbols have no internal structure (and are thus too ‘blocky’ to support genuine intelligence) and also embody the top-down mindset common to all traditional systems. As noted above, bottom-up approaches are required in order to generate true intelligence and AGI. Traditional symbols have none of the special features, properties, or abilities that atoms do and are thus insufficient for AGI.

Response to Dreyfus / Fjelland View That AGI Cannot Be Realized

In the June 2020 issue of Nature Humanities & Social Sciences Communications, Fjelland notes various difficulties that have historically arisen with respect to AGI, sees no way out, and consequently concludes that AGI “cannot in principle be realized” and is a “dead end” (p.3). He writes (p.2):

> *In 1976 Joseph Weizenbaum, at that time professor of informatics at MIT and the creator of the famous program Eliza, published the book Computer Power and Human Reason (Weizenbaum, 1976). As the title indicates, he made a distinction between computer power and human reason. Computer power is, in today’s terminology, the ability to use algorithms at a tremendous speed, which is ANI [Artificial Narrow Intelligence/Weak AI]. Computer power will never develop into human reason, because the two are fundamentally different. “Human reason” would comprise Aristotle’s prudence and wisdom. Prudence is the ability to make right decisions in concrete situations, and wisdom is the ability to see the whole.*
>
> *These abilities are not algorithmic, and therefore, computer power cannot—and should not—replace human reason. The mathematician Roger Penrose a few years later wrote two major books where he showed that human thinking is basically not algorithmic (Penrose, 1989, 1994).*

Fjelland acknowledges the critical importance of causality but, because in his view computers “cannot intervene in the world” and “do not grow up, belong to a culture, and act in the world”, they will therefore never be able to handle causality nor “acquire human-like intelligence”. (p.3) Fjelland further cites Dreyfus, who believed that “computers, who have no body, no childhood and no cultural practice, could not acquire intelligence at all” and that true AI is fundamentally impossible because an important part of human knowledge is tacit and therefore cannot be “articulated and implemented in a computer program” (p.2). Related arguments have also been made within the embodiment literature, especially with respect to linguistics and mathematics.

> *According to [Plato], a minimum requirement for something to be regarded as knowledge is that it can be formulated explicitly. Western philosophy has by and large followed [an incorrectly reductionist view of] Plato and only accepted propositional knowledge as real knowledge. An exception is what Dreyfus called the “anti-philosophers” Merleau-Ponty, Heidegger, and Wittgenstein. He also referred to the scientist and philosopher Michael Polanyi, [who] introduced the expression tacit knowledge. Most of the knowledge we apply in everyday life is tacit. In fact, we do not know which rules we apply when we perform a task. Polanyi used swimming and bicycle riding as examples. Very few swimmers know that what keeps them afloat is how they regulate their respiration: When they breathe out, they do not empty their lungs, and when they breathe in, they inflate their lungs more than normal.*
> *(p.3)*

The system presented here wholly overcomes each of these critiques. To begin, the unlimited causal and semantic nuance inherent in atoms meets the world exactly where it is and on its own terms, without simplification. Fjelland notes that “[m]ost of the knowledge we apply in everyday life is tacit.” (p.3) Tacit knowledge is indeed critical; our bottom-up, atomic approach is the first capable of properly accessing, storing, and applying it. Tacit (and, we argue, all true knowledge) is not directly accessible to human consciousness, but the tools and protocols embedded in this AGI enable us to unearth it, properly store It, prove its correctness, and use it to deliver AGI. Our bottom-up approach generates knowledge bases that directly match the semantic, systemic, and causal structure of the real world and that readily enable simulation of the human mind. (cf. Olsher and Toh 2013, Olsher 2012c ,2013c, 2013b, others) The system is able to represent commonsense knowledge (required in order to generate intelligence) as well as natural language semantics/meaning and syntax and can integrate these with other intelligent phenomena.

All of this very much places our system 'in the world' and, ironically, does so far better than humans, who can only interact with the world through consciousness (which is itself profoundly biased), models, perceptions, and unconscious belief systems. It actively rejects models and simplifications and is therefore able to see more clearly than humans. It is also capable of empathy far superior to, and can consequently be far more moral than, human beings. It is able to move itself towards what is good and autonomously impose limits on itself whenever morally appropriate.

As has been proven for the US Government and in previous work, our system readily and accurately represents and simulates culture, psychology, and the human mind, giving it an innate ability to understand people and see the world from diverse points of view. It deeply understands the real world and its courses of action are deeply grounded in reality. The system therefore does indeed "grow up, belong to a culture, and act in the world" and is able to meet Fjelland's criterion for achieving human-like intelligence. In addition, this AGI achieves what Weizenbaum terms "human reason" far better than humans - its capabilities for prudence and wisdom (Weizenbaum's "ability to see the whole") manifestly exceed those of humans and are achieved in a provably correct, entirely transparent, and provably unbiased manner humans cannot match.

Fjelland's thesis suggests that a core reason that it is essential for computers to be "in the world" is that without this they can have no understanding (p.6). As we saw above, understanding is indeed fundamental to AGI; this system is able to leverage its deep grounding in the real world to generate understanding in all senses of the word.

Fjelland writes that "we do not know which rules we apply when we perform a task" (p.3) and, here, we see echoes of why AGI has been so difficult to achieve – ***there are, in fact, no rules***. No rule-based system could achieve any aspect of intelligence, because rules are anathema to autonomous adaptation at runtime, which all genuinely intelligent systems *must* do, and they parasitize human intelligence, which any genuinely intelligent system *cannot* do. Fjelland suggests that the cost of moving away from rules is loss of explicitness, but this is clearly not the case – atoms are completely explicit yet fully achieve everything previous thinkers have attempted to achieve via rules. We *can* have our cake.

Fjelland observes that traditional systems are brittle and cannot handle change, which is certainly the case (see e.g. Olsher (2014) and Mueller (2006) for further exposition here). He discusses the extent to which Big Data has exploded, but, as we discuss in great depth above, data is not 'the new oil' – far from it in fact. Data is of no use in solving real-world problems because the best we can do with it is collect correlations, which are insufficient to generate intelligence. Correlations tells us what happened in the past, but if we don't know *why* we won't be able to extrapolate this into the future.

Atoms provide the necessary response to Pearl and Mackenzie's query as quoted in Fjelland: "How can machines (and people) represent causal knowledge in a way that would enable them to access the necessary information swiftly, answer questions correctly, and do it with ease, as a three-year-old child can?".

Fjelland (p.5) cites Hume's insight that "the distinguishing mark of a causal relationship is a 100% [connection] between cause and effect." (p.5) Traditionally, this has not been achievable because the world is too complex and shot through with too many deep contextual interactions that humans cannot see nor understand. **But atoms are small enough to represent reality at a level where *there is 100% correlation between cause and effect.* This insight is essential.**

Atoms convert *long* causal chains (which are indeed intractable in many ways) into small-component causal chains, *in which these issues no longer exist.*

Thus, genuine, proper causal knowledge *can* indeed be formulated explicitly, just not in traditional forms. True knowledge is therefore *atomic*, not propositional, in nature.

Traditional Approaches Can Never Reach AGI

In Fjelland's view, the belief that AGI can be realized is dangerous because it leads scholars to envision an impossible future. As we show here, AGI is indeed achievable, but *the notion that it can be reached from traditional AI* is exactly what Fjelland worries about - this belief pushes billions of dollars of investment towards projects that cannot possibly succeed and gaslights hopeful technologists into the belief that they can achieve the impossible 'if only they push a little bit harder'. An inescapable corollary of the present work is that traditional AI has hit its ceiling; the sooner we recognize this the faster we will be able to fully realize the benefits of AGI for humanity.

3. The Olsher Test - First Proper Evaluation Method For AGI

Previous scholarship, including but not limited to the Turing Test, has not been able to offer a proper test for AGI. There is a need for a fair and proper test, as stringent as possible, with firm theoretical grounding. We now provide that here.

The Olsher Test is the first proper evaluation methodology for AGI. It is deeply rooted in the theoretical and practical frameworks provided above and **embodies a 'kitchen sink' principle wherein *all potential requirements have been included in the Test*. *It is therefore as strict as it can possibly be.*** As has been shown above, via other work, and through practical application, ***the AGI presented here meets the Olsher Test in all respects***.

The Olsher Test's proof mechanism employs the strongest possible means - *proof by property*. **The Test is strong enough to provide results that are *general – that is, guaranteed to hold for all future cases*. *No traditional AI technology is strong enough to support this*.** Only weak evaluation methods may be used with traditional technologies because they are statistical in nature and are blind to context and change; past performance is therefore no guarantee of future success. They may work, or not, depending on the conditions experienced on any given day and may therefore only be evaluated application-by-application and instance-by-instance. Because nothing can be proved about them in the general case, traditional systems cannot possess generality, which must be proven of systems as wholes and cannot be inferred from instances.

Satisfying the Olsher Test in all respects is sufficient to definitively prove that AGI, Strong AI, and true generality have been achieved. If an AGI is able to pass the Test it is assured that it will work for any future problem; faith is no longer required. Showing that the Olsher Test has been met for a system overall is therefore far more powerful than demonstrations of individual applications.

The Turing Test Is Insufficient for AGI

Today's Turing Test ultimately centers around human deception, but is focused on the wrong things, and is unnecessarily capricious, from an AGI perspective. There are existing systems which claim to have met the Turing Test but are clearly not AGI. It is easy to fool people when they don't understand what is hard about AGI and don't know how to evaluate it. To pass, it is sufficient to merely regurgitate texts written by humans, but so doing represents no advance in AI.

If we consider at a minimum the multiple views of Turing's work set forth in Fjelland (2020), we can see that the Olsher Test represents what Turing ultimately sought but did not possess the theoretical foundations to provide – a genuine test for AGI that properly considers AGI fundamentals and directly answers the core questions of whether or not intelligence and/or AGI have actually been achieved. The gold standard is to prove that proper fundamental properties and other elements hold, which is what we deliver here. In this way, we are able to take Turing where he ultimately hoped to go with his test and far beyond.

3.1. Definition Of The Olsher Test

The Olsher Test consists of the following elements, each of which must be satisfied in full for the test as a whole to be satisfied:

- core requirements,
- further required properties, characteristics, capabilities, and capacities, and
- necessary know-how, practices, and procedures.

All elements must be shown as holding in all cases, in a general manner not requiring further proof.

In the following sections we use the term *capability* to describe things systems must be proved to be capable of doing in the general sense and *property* for properties that must hold across all aspects of systems as wholes. *Capacities* are generalized statements that must be true of systems across the board.

In addition, the modifier 'first class' below means that a system is not only capable of doing and/or representing something, but is able to do so in a manner that is 'fluent' and 'natural' along the lines of Davis et al. (1993) as per the foregoing discussions.

Full support for the specific contents of the Olsher Test is readily found within the foregoing discussions and the definitions of intelligence given in the literature.

The concepts, basis, theory, design, structure, selection, organization, order, presentation, content choices, and so on of the Olsher Test, as well as the applicability, adaptation, application, etc. of the foregoing within the field of AGI and AI all represent original work of the author. In hopes of forestalling claims of definitional arbitrariness, however, two aspects of the extant scholarship – definitions of intelligence and the specific wordings of the elements making up those definitions - have been directly adapted into the Olsher Test as given within Legg and Hutter (2007). For the avoidance of doubt, however, all other elements remain original.

3.1.1. Core Requirements

The core requirements of the Olsher Test are met when all characteristics, requirements, properties, capacities, system capabilities, and all elements described in the sections above have been shown to properly and fully hold of a system as a whole without limitation and without caveat. Deep motivation for each element is provided within the foregoing discussions, and readers should readily be able to convince themselves that the core requirements themselves are in fact proper and correct.

3.1.2. Further Required Properties, Characteristics, Capabilities, and Capacities

While the vast majority of the items in this section are in fact corollaries of the core requirements, this is not always obvious. Presenting these directly, as we do here, assists in clarifying the implications of the above discussions and is helpful in ensuring that no gaps remain within the material. It also facilitates a deeper understanding of the concepts of *optimal properties* and AGI *'smells'*.

It is critical to note that the lists below are not intended to be exhaustive; rather, they are intended to provide a clear basis for understanding relevant implications and understanding what AGI requires and how our system meets those requirements.

In order to best facilitate understanding, the properties and characteristics below have been broken into two groups; the items in the first group would generally be expected to hold of *knowledge representations* and those in the second of *systems as wholes*.

Items Further Required of Knowledge Representations

- Full support for non-contextualized information enabling real-time adaptation, readaptation and nuanced real-world context sensitivity
    - Must enable avoidance of all pre-runtime adaptation
    - Fully support non-adapted information
- Universality – able to represent any information in any domain as is and without changes
    - Capable of matching any meaning, causality, and/or structure
- First-class holism
- Unlimited nuance
    - Capable of providing nuance necessary in order to enable runtime adaptation
- Flexibly define and store any type of information
    - Fully unbiased, causal, fully reusable across contexts
    - Physical systems and connections to external systems can be readily represented
- Cross-domain integration – must be possible to reason seamlessly between disparate domains and apply common techniques regardless of source, domain of information
    - Full cross-domain fusion, including but not limited to semantics, causal and other structure, and properties
- Capable of representing information and causal chains at level of analysis small enough to enable properties to be proved
- First-class
    - Causality
    - Meaning
    - Nuanced general reasoning
    - Nuanced general simulation

- Maximum inference – must provably support all potential inferences achievable from knowledge bases (otherwise there will be certain things the system will not be able to 'think')
- Monotonic performance under knowledge addition – it must be possible to add new knowledge to the knowledge base without disrupting the current performance and/or provable properties of the system
- Noise resistance – it must be possible in some instances to accept some proportion of noise in the knowledge base without damaging system performance and/or provable properties
- Tractability – system operation must not include any operations that could become intractable
- Lack of search – a system must never depend on search and/or any other operation that may or may not succeed, be optimal, and/or potentially be intractable or not provably correct
- Enable information and causality to be represented at a low-enough level that transparency and all other relevant properties become feasible
- Provide full support for context, including but not limited to non-adapted representation
- Represent all forms of information in wholly unbiased fashion
- Enable all influences to be represented such that they may be holistically taken into account
- Faithfully represent implicit systems and semantics
- Enable proving of correctness in all respects and with respect to information, causality, and any other relevant aspects
- Enable information to be built without bias
- Fully transparent in all aspects
    - Readily amenable to human understanding
- Meet elegance and property floors and avoid smells
- Enable full support for all required system properties given above

Items Further Required of Systems As Wholes (including design properties)

- Able to compute and/or complete any intelligent task in real time and in view of context obtaining at that time (never limited to what designer foresaw)
- Deep, nuanced, complete understanding across all domains
- Retain all nuance, rather than requiring the simplification of thinking by e.g. removing detail
- Process exact states of the world in context
    - Unconstrained reality
- Fundamentally causal in all respects
    - Full causal awareness in all areas
- Maximal elegance in all respects
- Bottom-up understanding
- Full, deep capability, sensitivity, and support with respect to nuance across all areas
- Full, deep capability, sensitivity, and support with respect to context across all areas
- Incorporate nuance and context across all areas
- Fundamentally semantic in all respects
- Holistically take all influences into account across all areas
- Full support for adaptation and re-adaptation, including but not limited to real-time contextually-sensitive adaptation
- Provide specific solution to exact problem faced in context (vs. highly ambiguous results requiring further human effort in order to realize/make useful)
- No use of rules in any aspect
- Allow computer to discover what matters in context as opposed to human supposition
- No simplifying assumptions
- All system aspects meet elegance floor
- No AGI 'smells' (as described above, including but not limited to any use of training data, statistics, models)
- Complete transparency readily amenable to human understanding in all aspects
- Verifiability in all aspects

  It must be possible to verify correctness of all aspects simply by inspection. All reasoning, simulation, and all other aspects of system operation must be completely transparent and capable of complete, thorough verification.
- Complete avoidance of bias across all aspects

- Ability to explain self in all aspects, demonstrate why system is correct, answer human queries, and prove the correctness of all of the foregoing
- All intelligence must reside in computer itself, not at all in the human
- Computer must never borrow human intelligence or consciousness
- Computer must have its own independent consciousness
- No ‘parasitizing’ of human intelligence
  It must be possible to fully and transparently validate that no human beliefs or biases have been embedded within a system.
- Real-time contextual adaptation and reconstrual
- System must be autonomously, provably capable of safely and correctly reasoning, explaining, simulating, exercising morality, and exercising self-control
- Real-world usefulness (as defined above)
- Capacity to detect change and enable adaptation to it
- Strong generality (as defined above)
  - Including but not limited to seamless reasoning, simulation, and semantic and causal connections across any and all domains
- Faith never required
- Provably unbiased
- Provable safety, including all necessary understanding, ability to provably compute morality, and self-control
- Provably trustable
- Provably correct in all respects
- Provably unbiased
- Prove that bias is prevented from entering system in all areas
- Provably pro-human
- Provably safe
  - Provably correct and complete full moral simulation capability
  - Provably control itself within human-given limits
  - Provable ability to understand effects of potential actions and exercise transparent, provably correct self-control with respect to these
  - Ability to know when to ask for permission, to act responsibly in the meantime, and to explain to a human the costs and benefits of proposed choices
- Prove future actions will have all proper and necessary properties present
- Prove, by properties, that system will work on all future problems regardless of domain
- Provably independent consciousness
- Solve exact problem faced - adapt exactly to, and provide exact solution exactly for, problem being solved, never generic
  - No use of proxy problems
- Holistically take all influences into account
- All tools and requirements in place to properly perform runtime re-construal
- Capable of adapting Strong AI to any potential problem
- Deliver AI strength across all areas and aspects
- Provide elegant, optimal operations in log linear/non-exponential time
- Policies, procedures, and mitigations exist and have been proven to the greatest possible extent in all ‘difficult’ areas (such as creating knowledge without bias)
- Meet elegance and property floors and avoid smells
- Enable full support for all required system properties given above

Further Required Capabilities
Items in this section have been grouped visually for ease of reference.

- Understand
- Understand complex ideas
- Understand complex situations
- Understand and deal with new situations
- Understand principles, truths, facts or meanings, acquire knowledge, and apply these in practice

- Autonomously use, adapt information in context

- Simulate (with all ideal properties)
- Imagine
- Capability to project into the future

- Meet novel situations, or to learn to do so, by new adaptive responses
- Adapt to dynamic task demands in context and in view of the situation
- Adapt effectively to the total environment, including by changing self, environment, or finding new environment
- Adapt to new contexts, environments, and situations
- Capacity to reorganize behavior patterns so as to act more effectively/appropriately, esp. In novel situations

- Act purposefully
- Achieve complex goals in complex environments

- Handle/deal with/solve problems
- Handle/deal with/solve novel problems/problems the system and/or its designers never saw before

- Store, use, apply knowledge and experience in context, situation, goal aware manner

- Deal effectively with the environment
- Solve hard problems
- Have, use, benefit from knowledge of the world

- The resultant of the process of acquiring, storing in memory, retrieving, combining, comparing, and using in new contexts information and conceptual skills
- Mechanism by which the effects of complexity of stimuli are brought together and given a somewhat unified effect
- Understand and profit from experience
- Aptitude in grasping truths, relationships, facts, meanings, etc.

- Acquire and apply knowledge
- Acquire and apply skills

- Perform contextual adaptation and re-adaptation
- Avoid early adaptation across all aspects

- Judge
- Choose/Select

- Effective use of concepts and symbols in dealing with a problem to be solved
- Perform an operation on a specific type of content to produce a particular product

- Overcome obstacles by taking thought
- Reason/engage in various forms of reasoning/skilled use of reason
- Form, use, apply, adjust, accurately represent, reason with concepts
- Apply knowledge to manipulate environment or think abstractly
- Have opinions based on reason
- Think, including abstractly and about abstract things, with all desirable properties
- Exercise good judgment
- Plan

- Solve problems, or create products, that are valued within one or more cultural settings

- Comprehend ideas
- Comprehend language

- Learn, including from experience
- Learn facts and skills and apply them, especially when this ability is highly developed

- Adapt to any changing context, circumstances
- Adapt in contextually-sensitive way
- Think/Reason (transparently, fully unbiased, provably correct, nuanced, causal, context- and problem-aware, adaptable)
- Explain in depth exactly why system and outputs are correct
- Provably simulate the human mind
- Provably compute moral consequences and judgments
- Make provably correct nuanced moral understandings, decisions
- Deliver provably correct, verifiable, transparent reasoning usable across all problem types
- Deliver general simulator
- Deliver general problem solver
- Understand all forms of information in unbiased fashion
- Fuse all forms of information in unbiased fashion
- Understand contextual implications of all forms of information in unbiased fashion
- Simulate human mind in accurate, unbiased fashion
- Deliver general reasoner (tractable – i.e. without any search, O(log n) time in size of knowledge base, as ours does)
- Predict effects of own actions
- Define, understand, represent, simulate culture and psychology
- Generate output relevant to any potential problem
- Represent, provably reason about, simulate general complex systems
- Represent, provably reason about, simulate general cross-domain systems
- In provably fully accurate, verifiable, context- and problem-aware manner:
    - Flexibly define and store any type of information (fully unbiased, causal, fully reusable across contexts
    - Control self within human-given limits
    - Incorporate nuance and context across all areas
    - Holistically take all influences into account across all areas

Further Required Capacities
Items in this section have been grouped visually for ease of reference.

- Possession of general factor that runs through all types of performance (and throughout fundamental system design)
- Capacity for system to undertake activities that are characterized by (1) difficulty, (2) complexity, (3) abstractness, (4) economy, (5) adaptedness to goal, (6) social value, and (7) the emergence of originals, and to maintain such activities under conditions that demand a concentration of energy and a resistance to emotional forces
- Capacity for adaptation to relevant environment under insufficient knowledge and resources
- Capacity to act appropriately in uncertain environments, including where appropriate action is that which increases the probability of success and success is achievement of subgoals that support the system's ultimate goal
- Demonstrate intelligence [as] part of the internal environment that shows through at the interface between person and external environment as a function of cognitive task demands
- Capacity to "process information properly in a complex environment, including when criteria of properness are not predefined and hence not available beforehand and when these are acquired as a result of information processing"
- Successful (i.e., goal-achieving) performance of the system in a complicated environment

- Quickness, range, flexibility in association, facility and imagination, span of attention, quickness and alertness in response

- The power to rapidly find adequate solutions in what appears a priori to observers to be an immense search space **(**but while never employing search, which is impermissible in AGI systems – see above*)*

- The capacity to acquire capacity
- Capacity for:
  - Knowledge
  - Association
  - Memory
  - Sensory capacity, capacity for perceptual recognition

- That faculty of mind by which order is perceived in a situation previously considered disordered
- 'Clean' cognition ((fully transparent, provably correct causal reasoning and simulation)
- The computational part of the ability to achieve goals in the world
- Cognitive ability
- The general mental ability involved in calculating, reasoning, perceiving relationships and analogies, learning quickly, storing and retrieving information, using language fluently, classifying, generalizing, and adjusting to new situations
- Capable of doing what is appropriate for circumstances and goal
- Flexible in view of changing environments and changing goals
- Sensitivity to local environment
- Capable of making appropriate choices given perceptual limitations and finite computation
- The ability to perform tests or tasks, involving the grasping of relationships, the degree of intelligence being proportional to the complexity, or the abstractness, or both, of the relationships
- The ability to plan and structure one's behavior with an end in view
- Awareness of the relevance of self behaviour to objectives
- The ability to use optimally limited resources - including time - to achieve goals
- Ability to achieve goals in a wide range of environments
- Doing well at a broad range of tasks

- That combination of abilities required for survival and advancement within a particular culture
- Capacity of acting to advantage as social animal

- Ability to provide motivated direction for thoughts
- Capacity for abstraction
- Capacity for inhibition
- Capacity to inhibit instinct/previous beliefs/experience
- Capacity to realize internal changes in overt behavior

- Capacity for system improvement
- The essential, domain-independent skills necessary for acquiring a wide range of domain-specific knowledge
- Capacity to redefine in light of evidence/the results of imagination
- Capacity for learning from experience

3.1.3. Necessary Know-how, Practices, and Procedures

A final element of the Olsher Test consists of proving that a system and all relevant personnel possess all necessary and appropriate practical know-how and specialist expertise as described above.

References

Baumberger, Christoph. (2014) Types of understanding: Their nature and their relation to knowledge. Conceptus 40 (98).

Baumberger, Christoph. (2019) Explicating Objectual Understanding: Taking Degrees Seriously. Journal for General Philosophy of Science 50: 367–388.

Bostrom, Nick. (2014) Superintelligence: Paths, Dangers, Strategies. Oxford.

Brézillon, Patrick. (1996) Context in Human-Machine Problem Solving: A Survey. Research Report 96/29, LAFORIA.

Croft, William (2001), Radical Construction Grammar: Syntactic Theory in Typological Perspective. Oxford.

Davis, Randall, Shrobe, Howard, and Szolovits, Peter (1993) What Is a Knowledge Representation? AI Magazine 14 (1):17-33.

De Houwer, Jan, Barnes-Holmes, Dermot, and Moors, Agnes (2013) What is learning? On the nature and merits of a functional definition of learning. Psychonomic Bulletin and Review 20:631-642. Springer.

Egler, Miguel. (2021) Why understanding-why is contrastive. Synthese 199:6061-6083.

Fjelland, Ragnar. (2020) Why general artificial intelligence will not be realized. Nature Humanities and Social Science Communications 7, 10.

Gordon, Emma. (2012) Is there propositional understanding? Logos & Episteme, III, 2: 181-192.

Gordon, Emma (2023). Understanding in Epistemology. Internet Encyclopedia of Philosophy. ISSN 2161-0002, https://iep.utm.edu/, retrieved June 2023.

Grimm, Stephen. (2021) Understanding. Stanford Encyclopedia of Philosophy, https://plato.stanford.edu/archives/sum2021/entries/understanding/, retrieved June 2023.

Khalifa, Kareem. (2013) Is understanding explanatory or objectual? Synthese 190:1153-1171.

Legg, Shane, and Hutter, Marcus. (2007) A Collection of Definitions of Intelligence. Dalle Molle Institute for Artificial Intelligence Technical Report IDSIA-07-07, https://openresearch-repository.anu.edu.au/bitstream/1885/15009/1/Legg%20and%20Hutter%20A%20Collection%20of%20Definitions%20of%20Intelligence%202007.pdf, retrieved June 2023.

Mueller, Erik. (2006) Commonsense Reasoning. Elsevier Morgan Kaufmann.

Newell, Allen and Simon, Herbert. (1976) Computer Science as Empirical Inquiry: Symbols and Search. Communications of the ACM 19 (3): 113–126.

Olsher, Daniel. (2012) COGPARSE: Brain-Inspired Knowledge-Driven Full Semantics Parsing - Radical Construction Grammar, Categories, Knowledge-Based Parsing & Representation. International Conference on Advances in Brain Inspired Cognitive Systems (BICS 2012) LNAI 7366. Springer.

Olsher, Daniel. (2012b) Full spectrum opinion mining: integrating domain, syntactic and lexical knowledge. 2012 IEEE 12th International Conference on Data Mining Workshops (ICDM), Brussels, Belgium, pp. 693-700.

Olsher, Daniel. (2012c) Changing Discriminatory Norms Using Models of Conceptually-Mediated Cognition and Cultural Worldviews. In Naomi Miyake, David Peebles, Richard P. Cooper, ed, Proceedings of the 34th Annual Meeting of the Cognitive Science Society (CogSci 2012).

Olsher, Daniel. (2013) Cognitive-cultural simulation of local and host government perceptions in international emergencies. Proceedings, 3rd IEEE Global Humanitarian Technology Conference (GHTC 2013). 112-117.

Olsher, Daniel. (2013b) COGVIEW & INTELNET: Nuanced energy-based knowledge representation and integrated cognitive-conceptual framework for realistic culture, values, and concept-affected systems simulation. IEEE Symposium on Computational Intelligence for Human-like Intelligence (CIHLI) 82-91.

Olsher, Daniel. (2013c) AI Brain Sciences/Belief-Based Advertisements for Same-Sex Marriage Equality. Draft paper available via SSRN (https://ssrn.com/abstract=2234986).

Olsher, Daniel. (2014) Semantically-based priors and nuanced knowledge core for Big Data, Social AI, and language understanding. Neural Networks 58:131-147.

Olsher, Daniel. (2015) New Artificial Intelligence Tools for Deep Conflict Resolution and Humanitarian Response. Procedia Engineering 107: 282-292.

Olsher, Daniel and Toh, Heng Guan. (2013) Novel Methods for Energy-Based Cultural Modeling and Simulation: Why Eight Is Great in Chinese Culture. IEEE Symposium on Computational Intelligence for Human-like Intelligence (CIHLI). 74-81.

Parker, Jeanne, and Hollister, Debra (2014) The Cognitive Science Basis For Context, in Brézillon, Gonzalez, ed., Context in Computing: A Cross-Disciplinary Approach for Modeling the Real World, Springer.

Pritchard, Duncan. (2014) Knowledge and Understanding. In Fairweather, A. (ed.) Virtue Scientia. Virtue Epistemology and Philosophy of Science. Springer.

Waskan, Jonathan. (2003) Intrinsic cognitive models. Cognitive Science 27:259–283.