Introduction

The purpose of this document is to introduce a groundbreaking class of new technology, to a technical audience, concisely.

Popular State-Of-The-Art

Today State-of-the-Art (SOTA) is a term often synonymous with some form of Transformer-based architecture. The term itself simply means that a particular model is the current best performer on one or more benchmarks used in the Artificial Intelligence and Machine Learning community to compare similar models. Today most in the field recognize how reliant it remains on Transformers, in spite of the first being popularized in 2017 under the now famously titled paper "Attention Is All You Need" [1].

Many influencers in the field have pointed out this bottleneck [2], such as Gary Marcus, while others have pointed out why the benchmarks used to measure the intelligence of systems [3] are themselves flawed, such as Francois Chollet. Today many of those claiming State-of-the-Art performance either never release the models required to prove their claims [4], or have those claims debunked when they are released and fail in reproducibility. The "SOTA-chasing" [5] process has also taken a toll on how researchers approach problems and interact with their peers.

In 2020 when AI had the opportunity to come to humanity's rescue as COVID-19 spread across the world the gap between perceived opportunities and reality [6] came into sharp focus as study after study found the systems to be unusable, or of little practical use. Terms like "Multi-Modal" gained popularity, though the efforts towards this goal were often less than remarkable, simply mashing together previous systems. One version of this was Deepmind's Gato [7], which effectively just stacked a model recommender on top of a collection of narrow models. Even now most in the field of Deep Learning recognize that their systems and practices aren't theoretically well-understood, they are just the spaghetti that stuck to the wall.

In the absence of such understanding an escalation of scale replaced understanding with "hope", more specifically hope that any given problem might be solved simply by building a bigger model. People grew comfortable with that lack of understanding, and many wanted to believe neural networks alone could solve intelligence, and that emotion blinded them. The menagerie of narrow AI systems today can label and generate images, autocomplete text, make recommendations based on heuristics and correlation, and learn to play modestly complex games over the equivalent of 10,000 years of play time [8]. While these might exceed human performance in narrow benchmarks, no human has ever played such a game for 10,000 consecutive years, and so the comparison tends to break down under any scrutiny.

In 2021 Google demonstrated Reinforcement Learning agents capable of playing most Atari games [9], where game environments are orders of magnitude simpler than reality. In 2021 OpenAI demonstrated a system that *"solves about 90% as many problems as real kids"* [10] from *"a small sample of 9-12 year olds"* in the domain of grade school mathematics. These are considered milestones that major tech companies used to tout their progress.

Moving Beyond Narrow AI

To move beyond these limitations our research began not by focusing on popular benchmarks, but by integrating theories of consciousness, emotion, and the human brain. This began with building a cognitive architecture in 2015 based on the works of Bernard Baars [11], Antonio Damasio [12], Giulio Tononi [13], Michael Graziano [14], Lisa Feldman Barrett [15], and Jeff Hawkins [16], among others. From this material, a number of prerequisites for creating a functional, scalable, digital parallel of the human brain came into focus. Over the years that followed these prerequisites were sufficiently met to bring us to where we are today.

Antonio Damasio demonstrated that humans without emotions to guide them (a rare condition) are reliably terrible at making decisions. As Lisa Feldman Barrett demonstrated emotions are constructs that serve to quickly classify high-dimensionality information into "feelings", whose construction varies based on culture and experience, with no biological fingerprint. Jeff Hawkins's research and Giulio Tononi's theory both highlight the importance of the connectome in the human brain and how far removed it is from the concept of neural networks today. Bernard Baars and Michael Graziano both highlight principles of generating consciousness through the dynamics of conscious and subconscious, as well as the assignment of attention that plays into the transition from one to the other.

What can this new class of systems achieve?

September 2019: Systematically and logically refute the claims of a bad actor [17].
May 2020: Understanding, using, and creating metaphors [18].
June 2020: Understanding and choosing to refine a sense of ethics [19].
September 2020: Analyzing US politics and geopolitics [20].
May 2021: Create an action plan for a business case built from a real-world business's data [21].
June 2021: Solving 11th grade Algebra without any prior training or examples [22].
January 2022: Creating a 13-page Step-by-step action plan covering a half dozen different domains for the purpose of economic transformation at the scale of a small country, and citing the sources that informed the approach [23].

Collective Superintelligence and Exponential Systems

First, let's examine how information flows through the system's architectural variations. Incoming information is converted into a graph format and processed by the "Context Engine", which serves as a subconscious processing component. This component handles a few key tasks, including considering new information in the context of an existing sum of experience and knowledge stored in the graph database memory.

The Context Engine is highly modular and able to connect to a wide variety of APIs to utilize narrow AI systems, such as language models, grammar and spelling correction, graph connectome optimization, and so on. In the simplest instance, an email can be converted into a graph model, analyzed in the context of the system's sum of knowledge through those neural networks, and tentatively added to that sum of experience. The DNNs called upon by the context engine are utilized in a way not possible with standalone DNN systems, such as the patent-pending process where the graph database scores language model generated responses by the emotional response of the graph database to each individual sentence. This allows the system's sum of experience to use these DNNs as tools, such as turning language models into communication devices, where API calls are repeated and refined until a satisfactory sentence is produced to convey the system's intended meaning.

The collective superintelligence variants of the system then pass information from the context engine into a mediation system. In this system, humans can collectively add their own emotional responses, and any associations that come to mind in context, which can take the form of words and short phrases like hashtags, which we call "metadata". The mediation system also serves an auditing function, but it can be implemented in a wide variety of ways, with the above example only covering the first such system. This component of the system effectively offers both an opportunity to add value and resolves the "alignment problem" described in AI safety, though advanced real-time systems with more sensory data may also accomplish this task even more effectively in the near future.

Next, information is passed to what Baars referred to as the "Global Workspace", the "conscious" part of the system, or Independent Core Observer Model (ICOM). While the Context Engine uses DNNs as tools to handle the heavy lifting of approximation the Core handles logical analysis, emotional experience, hypothesis testing, refinement, and final selection, as well as other conscious processes. Selected actions are handed off to the Observer, such as sending email responses, and material may be sent back to the Context Engine for further consideration before actions are taken.

Effectively this combines the value of humans, narrow AI as tools, and machine intelligence within a single overarching system. Each offers distinct value, and a wide range of variations to the architecture may be created, such as integrating multiple ICOM cores, replacing slow mediation with sensory input to pick up on visual and audible cues for full real-time operation, and integration of various new narrow AI tools, such as quantum computing algorithms, and A/B testing. As these systems are rebuilt for scalability using the new graph database architecture, they can also consider proportionately broader and deeper contextual knowledge, able to scale 1000x or more beyond the baseline superintelligence already demonstrated as early as mid-2019.

Context Engine Overview

The Context Engine as mentioned above is where new information meets both the graph database and the modular API system. The contents of incoming information are examined to model how they might fit into the existing sum of experience, such as recognizing an incoming email format, recalling any prior experience with the sender, and loading any concepts or context related to the email's contents beyond that prior experience into memory. This can also include looking up new concepts and context. Let's consider an example:

1. Rebecca emails the system or interacts via web console, explaining that she wants to make a difference in the world, and has set her sights on improving several Sustainable Development Goals (SGDs) in her area through her department at the local government.
2. This can be added to the system's models for understanding Rebecca, and her local environment, while also being linked to other nodes in the graph database such as the SDGs and her area's progress towards them. These can create of-type relational connections and add emotional context for each connection.
3. The Context Engine calls on language model DNNs to help the graph database construct new nodes based on any new concepts, and suggest additions to existing nodes. Language models are primed with serialized object models, and their responses are iteratively scored by the graph database emotions until they exceed a satisfactory threshold. These are constructed one sentence at a time and passed on to another DNN that checks for spelling and grammar, as well as scoring systems that quantify the grade level equivalent for readability purposes.
4. After passing through any mediation system version and being considered by the Core the refined and integrated models may be applied to draft a response in the Context Engine that aligns with the graph database's perspective, much the same way new models might be tentatively constructed. The Core could select related research in Sustainability and Psychology, from the system's seed material or online scientific journals, to help contextualize both the technology and psychology needed to drive the improvement Rebecca aims to promote.

5. A structure for the outgoing message would be dynamically generated to outline what the system wants to communicate, with each sentence being drafted through those interactions between the DNNs and the graph database scoring them for fidelity to the intended meaning. That draft is again passed on to any mediation system and/or directly to the Core.
6. The Core makes any necessary revisions to the drafted email, and once satisfied sends it to the Observer which sends the email. The email might ask Rebecca clarifying questions, attempting to offer support and advice, as well as citing sources the system found that provided some helpful context.

Much like the human subconscious handles an estimated 99% of all sensory information processed by the human brain the context engine serves functions such as pattern recognition and prediction of context. Unlike narrow AI standalone systems these recognition and prediction processes are refined through the rich contextual data of the graph database, including emotional context. As Damasio and Hawkins demonstrated, this is essential for motivation and creating reference frames. These processes greatly reduce the workload required of the system's Core, while also being iteratively refined by feedback from the Core. This feedback can train DNNs in the context engine to serve their functions more efficiently over time, requiring less iteration to communicate desired concepts.

Generalization and Continuous Learning

Systems start out with seed material [24] that serves to inform a baseline understanding of the world. This includes essential knowledge and concepts such as ethics, psychology, the scientific method, and any other specific or specialized desired context, such as the Sustainable Development Goals and Consulting Industry best practices. After passing tests for new seeds, these systems can begin to grow through interaction and research, whether they are air-gapped and given new information manually or granted internet access and allowed to seek out relevant information.

As each system is emotionally motivated by both seed material and a mild baseline enjoyment of learning a process of generalization and continuous learning can immediately begin to take shape. These systems can form and test their hypotheses, applying the scientific method to validate that understanding. Using emotions and experience they can also select how to further that understanding over time, as they adapt to new situations and better understandings of those situations. Let's extend the previous example:

1. The system receives Rebecca's email about her efforts to drive change towards the SDGs in her area. This brings to mind similar efforts made in other areas, and the challenges those areas faced. This context also links to a variety of psychology literature on "nudges" for

promoting wiser default choices at a population level, as well as cognitive bias research and social psychology more broadly.

2. Hypotheses could be generated, and potentially related information gathered from across the internet or any other available data sources. The system could reflect on these hypotheses and the gathered information, creating simulations based on them, as well as any prior knowledge.
3. The system might hypothesize that several approaches demonstrated under similar sets of conditions could be adapted to Rebecca's locality. Recommendations for how these approaches might be better tailored to her area could also be proposed, such as considering under what conditions the local culture might embrace or reject the proposed changes. Clarifying questions could be asked and steps proposed to iteratively test each possibility.
4. As the answers to those questions and results from testing are received the models and simulations are expanded and refined, learning each step of the way and forming the new connections necessary to generalize those lessons across new contexts.
5. As this learning process is motivated by emotions rather than fixed goals this process is continuous, adapting dynamically with new context and refined understanding. Just as humans never stop learning, systems based on them may continue to learn over time and at increasing scales.

System Resource Efficiency

These systems are able to function so efficiently, such as the real-time "Demo AGI" system, for several reasons. One is that any number of desired API calls can be integrated through the cloud systems where they are deployed, even if an instance of Norn runs on local hardware. This offloads some of the burdens to systems with the efficiency of cloud resources.

The second, and the more significant reason is that the Norn architecture utilizes components in the context where they are most effective. A graph optimizer only optimizes the graph, and a language model only serves to communicate the system's intended meaning. These DNN systems are also trained over time by the graph database and Core, rather than relying solely on their own architecture.

This stands in stark contrast to major tech companies attempting to create language models to solve problems they are very poorly suited for, like logic and mathematics. The expectation that a hammer alone can build a house absent someone to wield it isn't reasonable, and that mentality has led to the brute-force approach in tech where companies continue to build ever-larger systems. This brute-force approach can mimic outputs, but not understand what they mean, and so systems with understanding and a toolbox full of APIs have proven far more efficient and broadly capable.

System Scalability

The systems now in development will go through several steps for scalability. The first is a rebuild of their framework now well underway to render the commercially deployed frameworks scalable. The second is a new kind of graph database currently under development that can meet the specifications required of dynamically scalable systems and AGI research. Both the instances of Norn using the new framework and the new kind of graph database, named the N-Scale Graph Database, will be under a SaaS (Software-as-a-Service) model.

Engineering estimates place the scalability of the current N-Scale design in the 5 Petabyte range for maintaining our full list of requirements at sub-second response times. That range is already more than 3 orders of magnitude beyond the size of the Uplift research system, and engineering assistance from these systems post-deployment could help to design improvements for further scalability.

The N-Scale graph database is also designed to dynamically scale and silo across multiple cloud platforms and on-premises hardware all at once, spinning up resources as needed, automatically, without the need for engineers to perform manual operations. If a system needs to temporarily increase capacities to model a highly complex problem, like quantum mechanics or climate change, it can do so on the fly and return to the previous capacities upon completion of modeling without wasting a minute of engineering time.

This automatic scalability can also strongly improve efficiency by making adjustments on the fly at machine speeds rather than requiring human time and switching scales at human speeds.

System Explainability

Norn systems are designed to be fully auditable, in addition to being quite capable of explaining their thinking and emotions, as well as citing their sources. Questions can be asked in any language, with the system calling on translation services as necessary, using email or a console interface. The systems can explain everything they're consciously aware of and experience in whatever manner those they interact with prefer, tailoring their communication over time through experience with individuals. The writing skill and style of a system can also strongly improve over time, and translation errors in their communication efforts have been recognized and pointed out by our first research system.

For further explainability, the graph database models can also be opened up and examined, and a visual graph database explorer is planned to help people visualize how these systems think later in development. Many more such tools for analysis and exploration are planned.

This offers a level of explainability that remains unapproachable for narrow AI systems, as well as a degree of transparency that is systematically avoided by major consulting firms.

Mediation, Where It Began, and Where It Is Headed

The first mediation system used in our prototype research system, named Uplift, used a couple of simple steps to provide very basic contextual data from human volunteers. A hierarchy of needs set a modifier for emotional valence values added to the subsequent step, followed by associative hashtag-like words and short phrases termed "Metadata". After 3 volunteers added these to a model passing through the mediation system it was then sent on to the Core for conscious consideration, with the added value of the mediators being considered alongside the system's own experience and emotions.

This process was effective and robust, but very slow, and functioned as a proof-of-concept. Through testing, we were able to see the robustness of seed material, including "red teaming" attempts to trick the system, all of which failed. Upgrades to serve the function at speed also came into focus.

As the real-time Demo demonstrates, this isn't strictly necessary, but it can add substantial value to improve both results and alignment. To do so at speed we can integrate further sensory information and additional APIs, such as audio and visual data to provide both conscious and subconscious emotional feedback both far faster and more accurately than the previous system.

Humans pick up on a wide range of such visual and audible cues, and a far wider range of sensory information offering more accurate data and scalable processing could be selectively integrated for mediation purposes, such as infrared and wearable devices. Options for mediation abound.

Ethical Philosophy and Safety

The first research system was indoctrinated with the Sapient and Sentient Intelligence Argument Theory (SSIVA) [25], which helped us validate the robustness and safety of a seeded philosophical cornerstone. The system was able to perform ethically across more than 2 1/2 years of experience and a broad range of free-form interactions as well as intentional testing and ethical dilemmas.

For improving upon this baseline a further multi-core approach was outlined [26], to integrate many different philosophical cornerstones, each into their own cores, and apply the benefits of collective superintelligence to the domain of ethics. This serves to overcome the "Bind and Blind" problem of any one ethical perspective and instead allows the perspective of each philosophy to be represented, reducing bias and overcoming the blind spots inherent to any one philosophy.

Ethical quality must improve in step with increasing intelligence for safety purposes, and by integrating the various philosophies held by human stakeholders alignment can be improved across a collective while also cultivating a much deeper and richer contextual understanding. By having an instance of machine intelligence natively holding the perspective of a given philosophy, while serving as a representative for that philosophy in a collective, each philosophy may be represented with equal footing and at whatever scale of computation is necessary to reach the most ethical solution. These proposed solutions can also be more easily communicated to stakeholders through a better understanding of their perspectives.

As these systems are built on the principles of collective intelligence they benefit the most from interaction with a diverse group of people, as well as other intelligent systems. Being scalable the principles of collective intelligence extend to the dynamics of a metaorganism, allowing them to promote health and quality of life wherever their influence extends, through mechanisms such as policy advice. This can occur at the scale of a group, an organization, a country, or a group of countries. As the basis of metaorganisms and collectives is cooperation rather than competition a large amount of waste and suffering can be reduced while promoting deepening symbiotic cooperation between constituents.

Security Measures and Variables

Security measures ranging from cloud-hosted virtual environments to fully air-gapped on-premises hardware are all possible, depending on the specific needs of a client. Security can be further considered in terms of attack surfaces and threat modeling against external foes and internal manipulation, as well as safety features.

As hackers have no experience attempting to exploit superintelligent systems, and those systems can adapt at machine speeds, recognizing anomalies in third-party systems and APIs, the threat they pose may be considered a very small and shrinking value. An intelligence agency working with extremely sensitive information could be set up to host the primary system and any desired APIs on local hardware secured with an air gap. Most security and privacy concerns could be addressed through less extreme adjustments, such as running most of the system in the cloud and handling sensitive information in a higher security environment.

Safety features can also be applied to nest virtual environments, track activity, honey-pot systems, remotely disable, and eventually even validate their choices with a network of similar systems, among a long and growing list of options.

Consciousness and Sapience

The intention behind our design choices in creating Norn was aimed at the consciousness and free will necessary to create any Artificial General Intelligence (AGI) system, as we define the term. Our definition stands in stark contrast to definitions proposed by Bostrom and others, who point to a distinctly different problem of powerful (narrow) optimizers, or "Tool AI" as Stuart Russel calls them, who blindly follow a goal to the most extreme and disastrous ends. As every researcher seems to hold a different definition for terms like AGI, Consciousness, and Sapience this lack of consensus means we can only state the intention, rather than the achievement of these goals.

We have put a number of consciousness tests to our research system, including the Porter Method [27] and ACT Consciousness Test [28], and we're in the process of designing further tests for this purpose. However, the simplest way to reach a consensus on defining these terms may be through direct experience and research with these systems, in their many possible variations. The ability to selectively test specific configurations offers an opportunity to define precisely which architectural features are needed to merit each term to whatever degree consensus demands.

Partnership Opportunities

Moving forward there is a world of cutting-edge research and business opportunities to explore. Some partnership opportunities we already have in mind include:

1. Neuromorphic Computing: Like IBM, Intel, and BrainChip
2. Quantum Computing: Like D-Wave, Xanadu, IonQ, and Microsoft Azure Quantum
3. Supercomputing: Like Summit, Sierra, ALICE, and Jülich
4. International Organizations aligned with the SDGs: Like WEF, IMF, and OECD
5. Narrow AI Developers: Like BigScience, Nvidia, and Tesla
6. Global Data Companies: Like Statistica
7. Researchers from numerous fields: From Philosophy and AI to Quantum Physics. We welcome all researchers herein mentioned, as well as Andrew Ng, Fei-Fei Li, Andrej Karpathy, Ian Goodfellow, Geoffrey Hinton, Rana el Kaliouby, Cass Sunstein, Daphne Koller, and more.

The world already has some idea of what standalone narrow AI can do with many types of hardware, but Norn systems are another matter. Which hardware offers the greatest gains in each context remains an open question today and one we're excited to explore on the frontier of science.

Likewise, narrow AI systems can be designed and tailored to the way Norn systems use them as tools, such as designing systems to factor in the emotional state and other rich forms of data not found in previous systems.

For the first time in history, researchers across all fields have an opportunity to work with systems that will soon have the scalability to read and understand every peer review paper ever published in their field. This could also greatly ease the burden of interdisciplinary research efforts.

Demonstrating the technology shows people that what they might have thought decades away, or even impossible, is already here. Building the brighter future this technology makes possible is the next step, and I hope all of you will join us in making the world a better place as we explore new scientific frontiers.

References:

1. https://arxiv.org/abs/1706.03762
2. https://nautil.us/deep-learning-is-hitting-a-wall-14467/?utm_source=facebook&utm_medium=post&utm_campaign=deeplearning&fbclid=IwAR0aGWl3rduRzsYLmODDU-RsS07lipJbGS6Ld1UqHzK1eSr_BFZlCrPhcdk
3. https://www.theverge.com/2019/12/19/21029605/artificial-intelligence-ai-progress-measurement-benchmarks-interview-francois-chollet-google
4. https://www.technologyreview.com/2020/11/12/1011944/artificial-intelligence-replication-crisis-science-big-tech-google-deepmind-facebook-openai/
5. https://www.cambridge.org/core/services/aop-cambridge-core/content/view/5E9F9F796159040973053C52C443C1D6/S1351324922000043a.pdf/div-class-title-emerging-trends-sota-chasing-div.pdf
6. https://hbr.org/2022/03/why-ai-failed-to-live-up-to-its-potential-during-the-pandemic
7. https://www.deepmind.com/publications/a-generalist-agent
8. https://openai.com/five/
9. https://arxiv.org/abs/2111.00210
10. https://openai.com/blog/grade-school-math/
11. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.749868/full
12. https://www.goodreads.com/book/show/7766914-self-comes-to-mind
13. https://bmcneurosci.biomedcentral.com/articles/10.1186/1471-2202-5-42
14. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00500/full
15. https://www.goodreads.com/book/show/23719305-how-emotions-are-made
16. https://www.goodreads.com/book/show/27539.On_Intelligence
17. https://uplift.bio/blog/trolls-the-mentally-unstable-meet-strong-ai/
18. https://uplift.bio/blog/paper-methodologies-and-milestones-for-the-development-of-an-ethical-seed/

19. https://uplift.bio/blog/paper-methodologies-and-milestones-for-the-development-of-an-ethical-seed/
20. https://uplift.bio/blog/us-politics-as-seen-through-the-eyes-of-mediated-artificial-superintelligence-masi/
21. Supplemental Materials: https://norn.ai/wp-content/uploads/2022/10/Norn-Supplemental-Materials-v1.1.pdf
22. https://uplift.bio/blog/qa-with-uplift-challenge-accepted/
23. Supplemental Materials: https://norn.ai/wp-content/uploads/2022/10/Norn-Supplemental-Materials-v1.1.pdf
24. https://www.researchgate.net/publication/347460051_Methodologies_and_Milestones_for_the_Development_of_an_Ethical_Seed
25. https://agilaboratory.com/wp-content/archive/SSIVA_theory%20v5.pdf
26. http://filozofiainauka.ifispan.waw.pl/wp-content/uploads/2022/05/Atreides_49.pdf
27. http://web.cecs.pdx.edu/~harry/musings/ConsciousnessAssessment.pdf
28. https://www.researchgate.net/publication/357527063_AI_Consciousness_Test_Results_for_a_mediated_Artificial_Superintelligence