2025

# The future of visual intelligence: AI Vision through the looking glass

By Jeremy Michaels

# Contents



# Prologue: Why this paper?

## The Crisinity: Living through remarkable and extraordinary times

In this 'post post-truth' era, we have moved beyond **Hypernormalisation**[1] into an age beset with unending crises. These create emerging opportunities, underpinned by transformative technological leapfrogs. Democratization of the internet; quantum theory; the transcendence of Artificial Intelligence (AI). The Age of Crisinity is here[2].

With the geopolitical, socio-economic, climate, and technological disruptions that we face, new opportunities for advancement are essential. The greatest shift in established systems, transfers of assets, and decentralization have long since been underway.

> I thought you could do anything with a camera that the eye could do, or the imagination could do. I didn't know that there were things you couldn't do. So, anything I could think up in my dreams, I attempted to photograph.
>
> **Orson Welles,**
> Filmmaker and scriptwriter

There is an emerging appreciation that we are today inexorably approaching what Ray Kurzweil called '**The Singularity'**. This is the hypothetical point when AI surpasses human intelligence, triggering unprecedented technological growth and disruption. More crucially, Kurzweil describes it as the point where 'humans transcend biology'.

## The Singularity: Arriving at true Artificial General Intelligence (AGI)

Whilst AI is truly of the zeitgeist, it traces its roots back to Alan Turing's '**Computing Machinery and Intelligence**' (1950) and Joseph Weizenbaum's '**ELIZA**' (1967). The next frontier is Artificial General Intelligence (AGI): AI capable of understanding, learning, and performing any intellectual task as well as or even better than a human being can.

Some believe in the same way as AI passed The Turing Test - silently and unnoticeable to the wider public - we are already at AGI. Yet proving it remains to achieve consensus. **DeepBlue** (1997) outperforming chess grandmasters, for example, may be one of the earliest examples of AI being comprehensible and reaching the attention of the public.

Judea Pearl's **groundbreaking** insights in Bayesian networks and causal reasoning led to his probabilistic approach (2011), one of the first to mathematize causal modelling in the empirical sciences. This, in turn, led to Kurzweil's '**Singularity Prophecy**', which is the theoretical point when intelligence becomes technically good enough to self-improve. It is this threshold where the exponential truly begins, and we move beyond AGI.

AGI may be achieved by 2030, or perhaps even sooner. With AGI, machine intelligence would not simply match but outstrip human cognition with recursive self-improvement and exponential knowledge growth. Put simply: humans infinitely enhanced via AI.

AI already exhibits broad intellectual abilities and performs high-level cognitive tasks like perception and context understanding. It might reason and plan in highly dynamic conditions, using its expertise in novel contexts. However, as per a **deep dive** into AGI by Nico Klingler (2025), it is essential to note that AGI still remains theoretical, for now.

With some claiming we have already achieved AGI, finding evidential proof is now the defining pursuit of our time. AGI is the next true frontier. Replicating and augmenting human vision – not just sight and cognition, but perception and understanding – is the most valuable aspect of AI, or what Fei-Fei Li, Professor of Computer Science at Stanford University, refers to within **Spatial Intelligence**.

---

[1] Hypernormalization (Definition): 'an accepted yet false societal system, that is so pervasive that whilst everyone knows it is in fact fake, people continue to act as if it is an objectively true and rational reality'

[2] The Age of Crisinity (Definition): 'the ability to find opportunities within the crises that we all face'

Yann LeCun, VP and Chief AI Scientist at Meta, outlined 'to achieve AGI we need to **train models on sensory inputs**' and 'of those, visual - or videos - are the most important. Effectively, 'humans see more data when measured in bits, and Large Language Models (LLMs) are primarily trained on text, carrying far less data'.

That being said, models have vastly improved their thinking by an order of magnitude over even just the last year.

Whilst machines remain unable to demonstrate the free will and consciousness that remain in a different realm to models, we retain a fundamental uniqueness in the universe we know. That remains untouched, but perhaps we are more predictable than we would like to believe? We may not be as special and complex as we think we are.

In this whitepaper we explore the path to Visual General Intelligence (VGI), and, by extension, AGI, and what it will mean for the future of humanity. We outline why the application layer matters most, what this unlocks in terms of innovation and benefits, and how this will not replace, but rather augment our progress and rapidly accelerate our evolution.

From a purely economic perspective, we conducted research backed up by **McKinsey & Company** and **PWC** estimates. Our analysis suggests that based on today's level of advancement, use, and potential alone, the current applications of Visual Intelligence (VI) will add anywhere between 2.5-4.2% to global GDP each year, within the next fifteen years. That could be as high as $5 trillion annually by 2040, from visual intelligence alone. We believe that is barely scratching the surface. The applications of VGI are effectively infinite..

It is possible to hypothesize that the value of VGI could surpass that even threefold by the time 'The Singularity' is reached. The implications - for the economy, science, medicine, longevity, innovation, exploration, and wider technology - are truly limitless.

> In a world overwhelmed by complexity, the systems we trust to manage it are today increasingly blind. What if we could give machines the power to truly see… not just to recognize, but to understand, anticipate, and act with real contextual intelligence?
>
> True visual intelligence is not just a technological leap: it is a new lens on the world, and one that could help us navigate crises, restore safety, boost productivity, and build trust across every layer of society. This whitepaper makes the case for why Visual General Intelligence will be the gateway to and quintessential proof point of Artificial General Intelligence… and why it may turn out to be the most important technology of our time.
>
> **Nico Klingler,**
> Co-Founder and Co-CEO @ viso.ai

## Visual general intelligence (VGI): The quintessential proof point of AGI

We see the most important proof point of AGI being the achievement of what we define as Visual General Intelligence (VGI): AI truly capable of understanding and interpreting visual information - across any domain – with, or surpassing, the agility, flexibility, and competence of human visual cognition, perception, and creativity.

VGI traces its roots back through early Computer Vision (CV) – also known as AI Vision – and now into the era of Visual Intelligence (VI): the most powerful application of AI today. True 'multi-modal grounding' is here: vision fully integrated into Large Language Models  (LLMs) for basic visual reasoning. 'Embodied agents' that navigate environments with learned visual memory and active control have also become established and scaled rapidly.

We are not far from true perception and visual understanding: systems processing and interpreting environments in real time with human accuracy. Imagine for a moment:

- VGI systems performing open-world visual reasoning and decision-making
- Vision-guided agents operating seamlessly in unstructured physical worlds
- Meta-perceptual reasoning, where systems understand what could or should be seen to guide perception

VGI is not only real, imminent, and powerful, but also offers hope in solving today's ongoing crises and the biggest challenges facing the human race for decades to come.

Visual Super Intelligence (VSI): perception, understanding, and simulation, with high-dimensional fidelity and abstract representation, is achievable. Systems reconstructing, simulating, and visualizing realities from solely mental input and with minimal data.

In an 'Age of Crisinity', we predict that Visual Intelligence is the major opportunity that can, in many ways, resolve and overcome many of the most pressing crises we face. Below we outline the 'why', 'how', and 'when' of VI. For now, let's begin with the 'what'.


 | 

viso.ai

# 1. Through the looking glass

" The future of AI is not just about making computers smarter but about making them see and understand the world like humans do.

**Andrej Karpathy,**
AI Researcher and co-founder of OpenAI

# 1. Through the looking glass

## In summary: Focus of this section

*This chapter introduces VGI as the next leap beyond traditional CV and VI, radically redefining how machines perceive and understand the world. VGI surpasses human visual cognition in depth, adaptability, and context awareness, marking a paradigm shift in AI.*

*It argues that VGI could serve as the most concrete and measurable proof point for AGI, offering tangible benchmarks through human-machine comparisons in visual comprehension, perception, and reasoning.*

VGI represents a journey 'through the looking glass' into an augmented reality: one where the traditional hierarchy of perception is fundamentally transformed. Just as Alice discovered a world where logic operated by different rules, we are entering an era where machines will surpass human visual cognition.

This is not merely true in terms of processing speed or data volume, but in the depth and sophistication of visual understanding itself. These systems will perceive patterns potentially invisible to human eyes, comprehend spatial relationships in dimensions we cannot fathom, and extract meaning from visual information with unsurpassed clarity.

This transition promises to reveal that what we considered the pinnacle of visual intelligence - human perception and understanding - was merely the beginning of a far more expansive visual consciousness. Machines will not simply see better than we can: they will redefine what it means to see, transforming our understanding of reality itself through their superhuman visual comprehension. Before we begin our journey through the looking glass and prove VGI, we wanted to clearly define the terminology we use.

## (A) Setting the scene: AI today

### (i) The here and now: Understanding Artificial Intelligence (AI) today

***Definition: Artificial Intelligence (AI)***

AI refers to systems that can perform tasks typically requiring human intelligence (learning, reasoning, problem-solving, and understanding language).

AI achieves this through Machine Learning (ML) techniques that enable computers to learn from data without explicit programming. Within ML, Deep Learning (DL) is a particularly powerful subset that uses neural networks with multiple layers to automatically discover complex patterns and representations in large datasets.

### *Definition: Computer Vision (CV)*

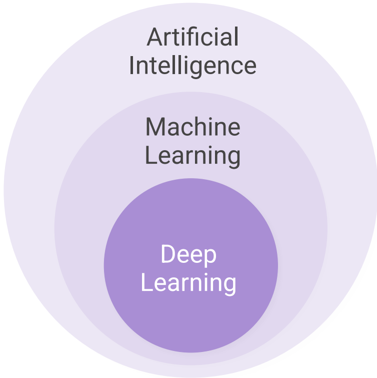
CV enables the interpretation, analysis, and understanding of information from images and videos, mimicking human vision, cognition, and perception.

### *Definition: Visual Intelligence (VI)*

VI is a shift from basic image processing and pattern recognition to more sophisticated systems, creating human-like visual comprehension through:

- Reasoning about visual content
- Understanding context and relationships between objects
- Making inferences about scenes, and
- Integrating visual understanding with abilities like language and decision-making

Developments in AI




## (ii) Looking to the future: The dawn of Visual General Intelligence

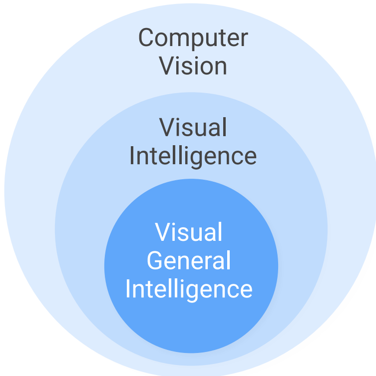
### *Definition: Artificial General Intelligence (AGI)*

AGI is a hypothetical form of AI that would possess human-level cognitive abilities across all domains, capable of understanding, learning, and applying intelligence to any task that a human can perform.

### *Definition: Spatial Intelligence (SI)*

SI is the cognitive ability to understand, visualize, and manipulate 3D relationships between objects and navigate through physical or conceptual spaces.

Emerging VI

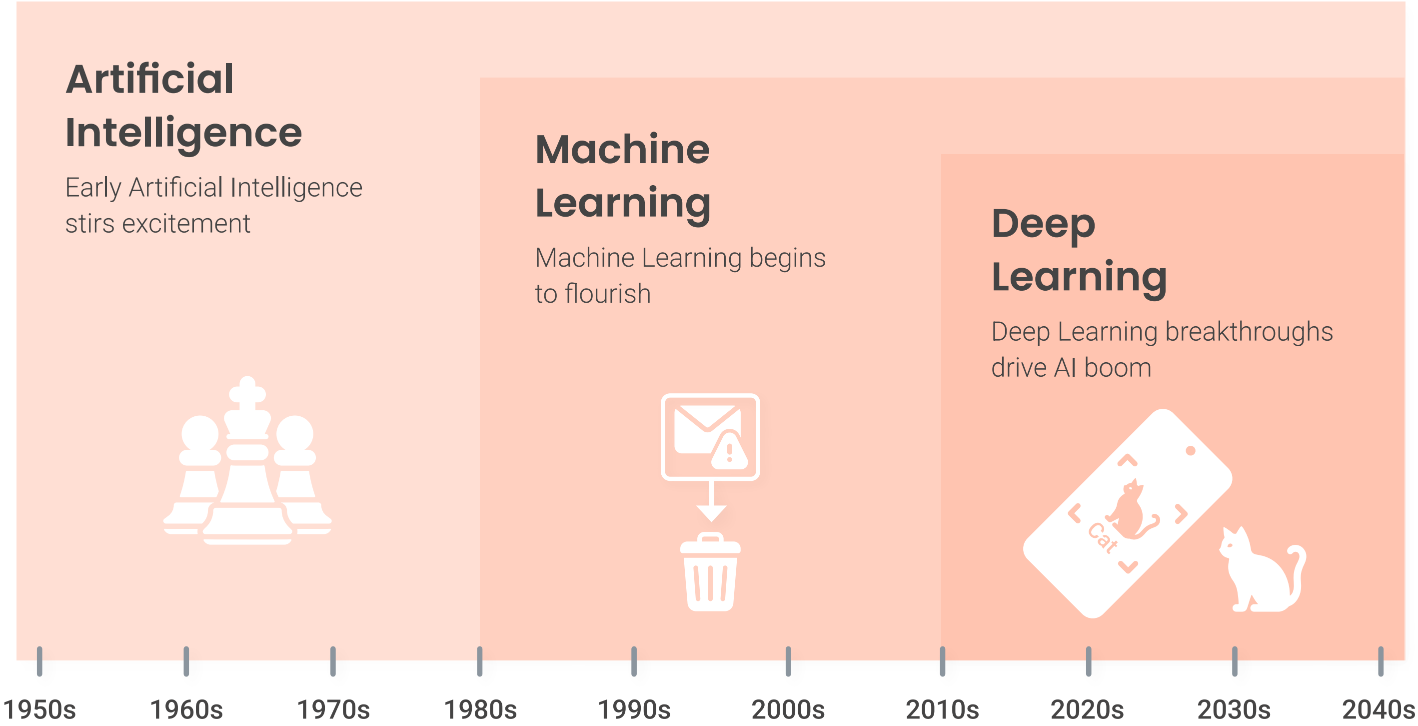



### *Definition: Visual General Intelligence (VGI)*

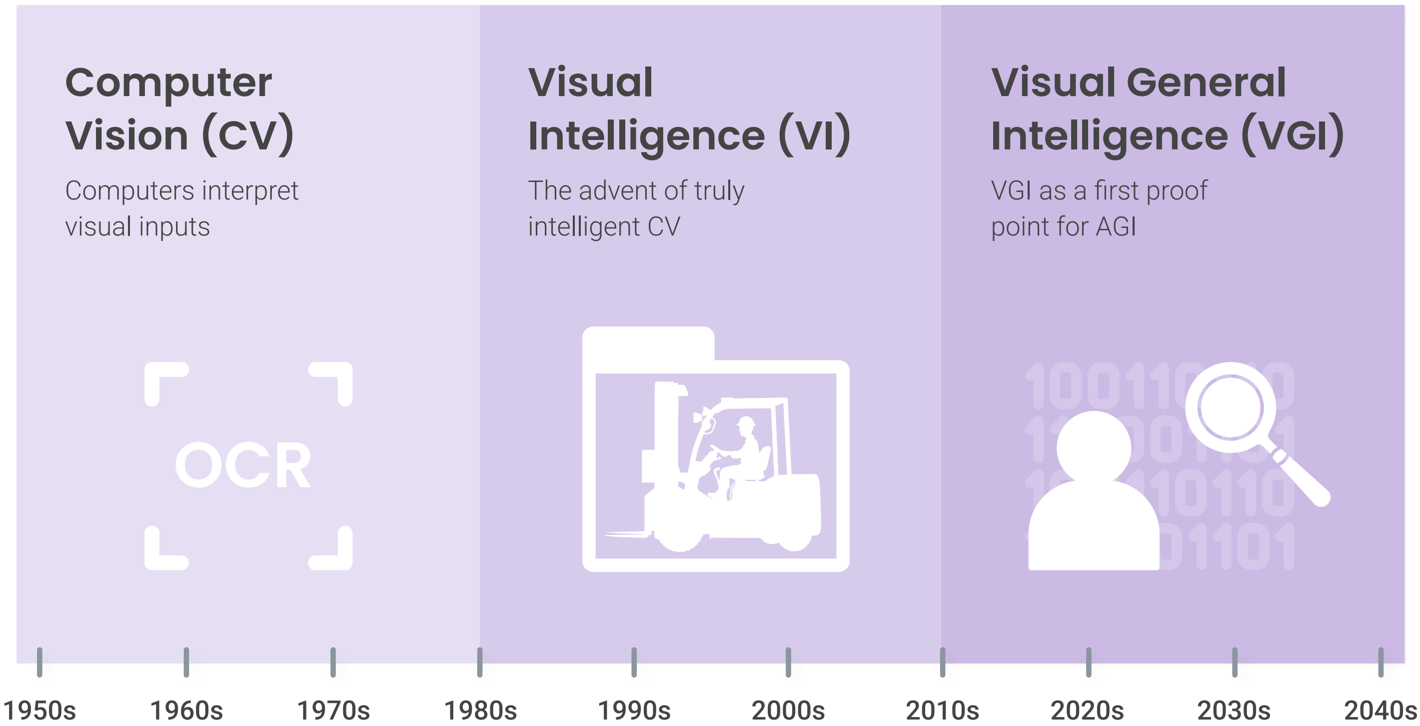
VGI is the capability of AI systems to understand and reason about visual information across all domains and contexts, with human-level comprehension, perception, and understanding. This enables the interpretation of any visual scene, inferences, and applying visual knowledge to solve diverse problems.

## The evolution of AI




## The emergence of VI



## (B) VGI proves AGI: Why Visual Intelligence matters most

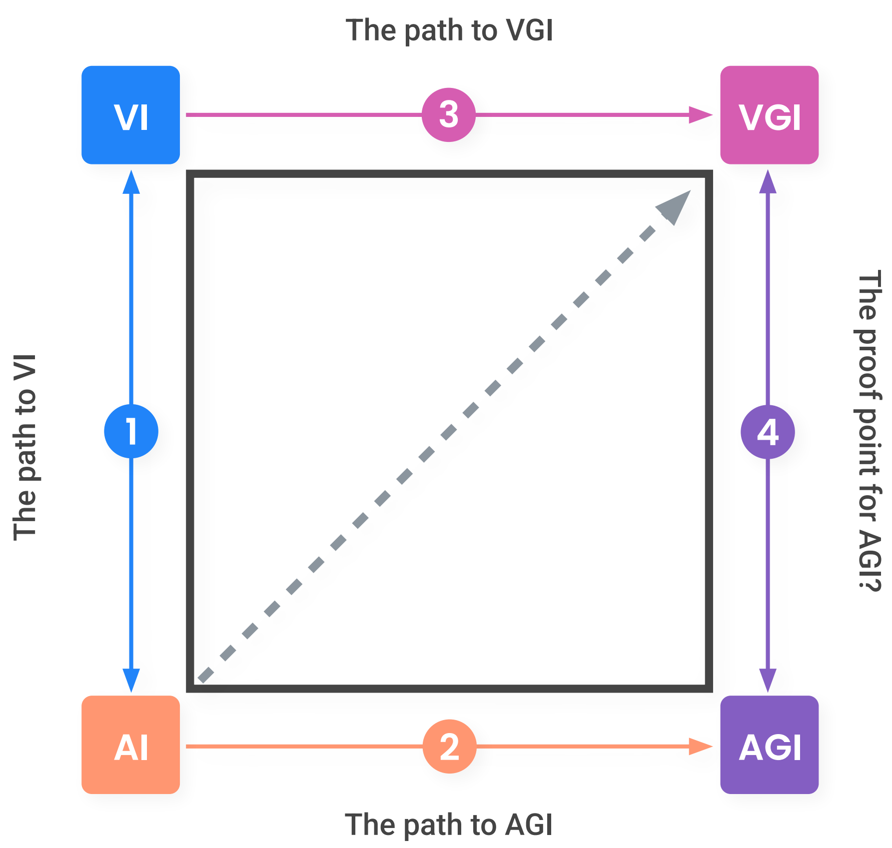
### VGI will be the most important proof point for AGI

AI has rapidly evolved and continues to do so. Breakthroughs, firstly in ML and then DL, have led us to where we are today. CV - which emerged from AI - is also evolving into what we consider to be true VI.

A potential proof point of AGI might be as simple as that it provides for a system or computer to be able to perform a task, entirely better than any human has ever done before (or could in the future). Proving this for a complex set of specific tasks or outcomes that are objectively quantifiable is far easier than applying it to general concepts.

In this regard, we believe it is interesting to consider: will VI – also on its own path to VGI - become the most important, reliable, and valid proof point of AGI?

In the case of VGI, if 20:20 vision and medically accepted metrics for human cognition (perception, awareness, recall, logic, and application) are both tangibly definable, how might VGI surpass that?  We believe that VGI could well be the most important proof point for proximity to AGI, and we set out below to detail why this is and how to prove it.




 | 
viso.ai

### Why do we believe this to be true?

- Vision is considered the most important of the five senses
- VI is the most complex and valuable version of all 'Sensory' or 'Spatial' AI
- VI is easiest to benchmark and measure, and hence provides objective proof

Yann LeCun, VP and Chief AI Scientist at Meta, explained that 'to achieve AGI we need to train models on sensory inputs' and that 'of those, visual - or videos - are the most important'. He asserted that 'humans see more data when measured in bits, and LLMs are primarily trained on text, carrying far less data'.

LeCun used the analogy that a four-year-old child with 2 million optical nerve fibers, carrying about 1 byte per second each of data, has seen more data than even the most advanced of LLMs across a similar time period. The challenge, therefore, is significant.

### When might VGI be achieved?

We believe that a leapfrog may well occur, analogous to that achieved by DeepSeek vs ChatGPT. However, with the pace of change so rapid, and definitions fluid, it is hard to predict. Current forecasts and estimates include that AGI (and potentially VGI first) or 'The Singularity' may be achievable in just a few short years.

A comprehensive analysis of 8,590 predictions indicates a median forecast for AGI **around 2040**. Notably, however, entrepreneurs tend to be more optimistic. Sam Altman (OpenAI CEO) envisions AGI emerging **by 2027**, emphasizing the need for responsible development to mitigate risks. Whereas, Demis Hassabis (Google DeepMind CEO) broadly concurs, anticipating AGI in 5-10 years, **between 2030 and 2035**. Ray Kurzweil has long predicted that AGI could be achieved faster than his original 2045 estimate, **as early as 2029**, based on the exponential growth of computing power.

### How might we know we have got there?

Within VGI benchmaking vision, understanding, and perception provides a powerful proof point for comparing human and machine visual intelligence. We believe that perception, understanding, and application matter most, proven by medically and scientifically robust methods, evaluation criteria, and metrics.

Assessing eyes, brain, and synapses vs cameras, systems, interfaces, and models, we could define tests comparing their relative 'hardware and software' and their capability. For example, **neural signals** distinguishing target from distractor images can diverge by ~150 ms, with correct categorizations in <300 ms - an ambitious latency target for VGI.

### And why does it matter?

If VGI can be proven to be the path to AGI, the implications for computer vision - both those in that sector and benefiting from the technology – would be significant for the global economy, tech sector, and AI.
For this, the application layer matters above all.

## (C) The starting gun: Market forces accelerating adoption

Several trends are converging to make VGI necessary and even inevitable.

### 1. The foundation model race: From lab to application

Open-source models led to a wave of innovation. Businesses can build on cutting-edge research, shifting focus from building models to effectively scaling and implementing.

### 2. Hardware democratization: Real-time inference everywhere

High-performance GPUs are today both more affordable and accessible. Companies no longer need to invest in massive data centers to run AI vision at scale. Edge devices and lightweight computing units make real-time inference feasible in any environment.

### 3. Data dependency and reduced annotation burden: Slashing timelines

Breaking data dependency and reducing data limits accelerates learning. Closed loop systems continuously learning and improving mean less annotation of data, generation of augmented data, and human-in-the-loop (HITL) feedback for user-led improvements.

Modern models require vastly less data to get better results (models generalize better). This is relevant as companies don't want to give up their data. Additionally, automated or assisted labelling makes the remaining annotation work quicker and easier.

Thanks to new training techniques, models can be taught using partially labeled or even unlabeled data. This dramatically reduces one of the most time-consuming steps in the computer vision pipeline, cutting months off project timelines.

> “ What excites me about the technology stack? We're talking about the convergence of edge computing, real-time inference, multimodal fusion, and embodied AI. About systems that process terabytes of visual data in milliseconds, distribute intelligence across millions of endpoints, and maintain context across time and space.
>
> The application layer isn't just important - it's where the magic happens, because VGI without applications is just expensive computer vision. VGI with the right application architecture... that's when we stop building software and start building digital senses.
>
> When machines can see and understand visual context globally, all systems become sensors, every interface, intelligent, and every application predictive. And that's when building reactive systems stops, and building anticipatory one begins.
>
> **Gerard Corrigan,**
> CTO @ viso.ai

## (D) Why the application layers matter most

In AI, attention is paid to model accuracy, benchmarking, and research breakthroughs. But true business value is in the intersection of AI and real-world operations. This is the 'application layer'. And this is where AI becomes more valuable and usable.

It is where insights from vision models feed into business systems, drive decisions, and automate action. A model may be 99% accurate, but unless integrated into workflows, configured for edge cases, and monitored post-deployment, it won't deliver results.

The application layer is no longer an afterthought: it is the beating heart of scalable, adaptive, and intelligent systems. Without it, even the most powerful models remain inert. With it, they become operable, generalizable, and valuable. In an era of model commoditization, the application layer defines competitive edge and shapes adoption.

If the data cannot be used outside of data science systems, then the business value is limited, as it remains in silos. The ability to build systems that extend beyond the data science and model layer unlocks the true value of and facilitates the leap to VGI.

### (i) Why it matters for the technology industry at large

The evolution from early CV to VI, and the leapfrog toward VGI, mark a pivotal inflection point in the development of AI. While foundation models and vision transformers push the frontier of perception, real-world impact depends on how models are applied.

Visual data offers the richest modality for understanding the world. Yet without structured application layers, this rich data is just inert potential. In the same way that the smartphone revolution was unlocked not by hardware alone, but by the app ecosystem (think iOS and Android), the application layer becomes essential for unlocking the power of VGI. It translates complex models into usable tools that businesses can interact with, configure, and derive insights from: all at scale.

Consider that Chat GPT was the first to introduce an application layer for text-based LLMs, even though several teams, including Meta AI, had this technology for years but did not manage to put it to use. New entrants to the sector that utilize available open weights, in combination with smart engineering, can challenge and outperform.

DeepSeek delivered comparable, and by some benchmarks, superior, performance at a fraction of the cost and development time of Chat GPT. This wasn't just about model architecture: it was about engineering efficiency, strategic focus, and rapid iteration.

Similarly, VGI may outpace legacy visual intelligence approaches not by out-modelling them alone, but by enabling faster, broader, and more modular applications through the right integration layer. It is this layer - connecting perception to action - that determines how quickly vision AI translates into real-world value.

## (ii) What the application layer means for Visual Intelligence

The application layer in the context of CV refers to orchestration systems connecting:

- Inputs (cameras, sensors)
- Vision models (e.g., object detection, segmentation, recognition)
- Processing/deployment tools (image transformation, edge computing)
- Output logic (APIs, storage, dashboards, notifications)

This is not about deploying models: it's about making VI operable, adaptable, and valuable in unpredictable environments. Key functions of the application layer include:

- Real-time configuration: Dynamic zones, thresholds, and workflows per site
- Edge deployment and orchestration: Efficiently running models across thousands of distributed devices
- Human-in-the-loop feedback: Integrating user corrections to improve model behavior and interact with outcomes
- Visual dashboards and insights: Bridging vision with business intelligence

> “ Training a model is a learning problem. Running it is an infrastructure problem. Impact is an integration problem: this is where most of the value is unlocked.
>
> **Gaudenz Boesch,**
> Co-Founder and Co-CEO @ viso.ai

Unlike model-centric platforms, what's needed is a full-stack platform that includes orchestration, security, data handling, and user interface logic.

## (iii) The transformative impact of VGI on global industries

We outline below a summary of the impact on key industries adopting AI Vision, the challenges they face, how they benefit from the application layer, and their VGI path.

### *A. Manufacturing: Producing better outcomes, faster*

- **Challenge**: Diverse environments, varied camera setups, changing products
- **Benefit**: Low-code tools to define custom logic (e.g., defect detection per line), edge deployment to reduce latency and bandwidth costs
- **VGI path**: Real-time adaptation to new SKUs, minimal retraining

### *B. Construction: Vision that scales with every stage of the build*

- **Challenge**: Dynamic environments, safety monitoring, harsh conditions
- **Benefit**: Remote configuration, site-specific rule setting, multi-angle monitoring
- **VGI path**: Adaptive learning for new waste types, contextual understanding of sorting requirements, movement, and routing

### *C. Waste management: Tackling contamination, sorting and recycling risks*

- **Challenge:** Contamination variability, sorting accuracy, operational efficiency
- **Benefit**: Automated classification rules, real-time contamination alerts, integration with facility management systems
- **VGI path**: Adaptive learning for new waste types, contextual understanding of sorting requirements, movement and routing

### (iv) Future use cases enabled by VGI through application layers

VGI unlocks new frontiers, but only when deployed through effective application layers:

#### *A. Adaptive agents: Autonomous perception and auto-tuning AI*

With embedded visual agents, applications could self-optimize via:

- Reconfiguring thresholds
- Suggesting detection categories
- Self-healing from edge failures

#### *B. General visual memory systems: Event and action understanding*

Applications may create conditions for episodic visual memory, allowing:

- Historical event querying (such as "when was the last forklift accident?")
- Long-term pattern discovery
- Cross-site comparative analysis

#### *C. Simulation and visual forecasting: Predictive vision models*

Imagine systems that can:

- Visualize future states ("where will this pallet be in 5 mins?")
- Generate alternate scenarios ("what would have happened if…?")
- Use generative vision models for synthetic training data


# 2. Ready?
# The next industrial revolution

> Every safety incident, supply chain disruption, quality failure, and downtime event happens in the visual world first. Your plant floor, warehouse, supply routes, safety protocols... all depend on human eyes catching problems before they occur.
>
> Eyes get tired and miss patterns. In complex operations, that's not just inefficient: it's dangerous. The leaders who understand this aren't just preparing for better monitoring systems: they position themselves to prevent problems that haven't happened yet.
>
> Imagine operations where equipment failures are predicted months in advance. Imagine supply chains that reroute themselves around disruptions before they occur. Imagine safety systems that identify risks in real-time and intervene before injuries.
>
> **Chrissie Jamieson,**
> Vice President Marketing @ viso.ai

## 2. Ready? The next industrial revolution

### In summary: Focus of this section

*Here, we explore how VGI will further transform operations across manufacturing, construction, and waste management through two key applications: Health and Safety and Lean/Overall Equipment Effectiveness (OEE).*

> If we want machines to think, we need to teach them to see.
>
> **Fei-Fei Li,**
> Professor of Computer Science
> at Stanford University

---

### (A) Health and Safety: The path to zero harm and zero downtime

#### 1. Manufacturing transformed: The eyes that never blink

VGI will revolutionize manufacturing safety by creating adaptive systems that learn from every incident and near-miss across global facilities. Unlike traditional safety monitoring that relies on rigid rule-based detection, VGI systems can understand contextual safety risks, predicting potential hazards before they occur.

Systems adapt to new equipment, changing production layouts, and evolving safety protocols without complete retraining: continuous protection as environments evolve.

Additional VGI-enabled HSE use cases could include:

- Predictive ergonomic analysis that monitors worker posture and movement patterns to prevent repetitive strain injuries
- Dynamic risk assessment that evaluates changing environmental conditions like temperature, humidity, and air quality in real-time
- Intelligent emergency response coordination that can automatically guide evacuation procedures and coordinate with emergency services based on visual analysis of incident severity and location

## 2. Construction: Hazard detection for ever-changing builds

In construction, VGI transforms safety management by providing comprehensive site-wide hazard detection that adapts to the constantly changing nature of construction environments. The system learns to recognize unsafe conditions across different project phases, weather conditions, and site configurations, while continuously updating its understanding of safety protocols as construction progresses.

This adaptive capability ensures that safety measures remain effective even as sites evolve from foundation work to finishing phases.

Enhanced VGI applications include:

- Fall risk prediction through advanced spatial analysis of worker positioning relative to edges and openings
- Heavy equipment collision avoidance that tracks machinery movements and predicts potential conflicts with personnel, and
- Automated compliance documentation that generates visual evidence of safety protocol adherence for regulatory reporting and insurance purposes

## 3. Waste management: Automate early-risk detection at scale

VGI enables proactive safety management in waste facilities by recognizing hazardous material handling violations and environmental safety breaches before they escalate into incidents. The system adapts to different waste streams, seasonal variations, and facility modifications while maintaining consistent safety standards.

This intelligence extends beyond simple detection to understanding the context of safety violations and their potential consequences.

Additional safety applications include:

- Toxic exposure monitoring that visually identifies potential chemical hazards and tracks worker exposure levels
- Fire risk assessment through thermal and visual analysis of combustible material accumulation, and
- Automated incident response that can immediately alert emergency services and guide evacuation procedures based on real-time hazard assessment

## (B) Lean and overall equipment effectiveness (OEE): continuous improvement

### 1. Manufacturing: predict, plan, and maximize throughput

VGI will revolutionize manufacturing safety by creating adaptive systems that learn from every incident and near-miss across global facilities. Unlike traditional safety monitoring that relies on rigid rule-based detection, VGI systems can understand contextual safety risks, predicting potential hazards before they occur.

Systems adapt to new equipment, changing production layouts, and evolving safety protocols without complete retraining: continuous protection as environments evolve.

Additional VGI-enabled HSE use cases could include:

- Predictive maintenance scheduling that optimizes repair timing based on actual equipment condition rather than fixed intervals
- Dynamic production planning that adjusts workflows in real-time based on demand fluctuations and resource availability, and
- Intelligent resource allocation that automatically redistributes personnel and materials to eliminate bottlenecks and maximize overall system efficiency

### 2. Construction: always-on productivity and equipment usage assessment

In construction, VGI enables lean project management by providing real-time visibility into resource utilization, workflow optimization, and progress tracking. The system adapts to different construction phases, weather conditions, and project complexities while maintaining focus on eliminating waste and maximizing productivity.

This enables making data-driven decisions: improving efficiency and quality outcomes.

Enhanced efficiency applications include:

- Automated progress tracking that compares actual construction progress against planned schedules and identifies optimization opportunities
- Resource optimization that ensures optimal allocation of personnel, equipment, and materials across project phases, and
- Quality-efficiency correlation analysis that identifies the optimal balance between construction speed and quality standards

### 3. Waste management: avoid shutdowns and optimize utilization

VGI transforms waste management: optimizing collection routes, sorting operations, and facility utilization via real-time analysis of waste streams and operation conditions.

The system adapts to seasonal variations, changing waste compositions, and facility modifications, optimizing for maximum efficiency and minimum environmental impact.

Additional efficiency applications include:

- Dynamic route optimization that adjusts collection schedules based on real-time bin fill levels and traffic conditions
- Automated facility layout optimization that reconfigures sorting and processing areas based on waste stream characteristics, and
- Energy efficiency monitoring that optimizes facility operations to minimize power consumption while maintaining processing capacity

# 3. The past does not dictate the future

> The brain has about 10^14 synapses and we only live for about 10^9 seconds. So we have a lot more parameters than we have seconds of existence.
>
> **Geoffery Hinton,**
> 2018 Turing Award winner for pioneering deep learning and neural networks

# 3. The past does not dictate the future

## In summary: Focus of this section

*This chapter presents VGI as a transformative leap from traditional CV and towards fully-adaptive, real-time, end-to-end VI systems. Unlike legacy wapproaches dependent on rigid rules or vast annotated datasets, VGI introduces flexibility, contextual awareness, and scalability for real-world deployment. Drawing on case studies like DeepSeek's cost-engineered breakthroughs and outlining the core architecture of VGI - including models, deployment, configuration, feedback, and analytics - it positions VGI as both a technological and strategic inflection point for AI vision's future.*

## (A) On the crest of a wave: The founder view

### The past does not dictate the future

- Not all vision-related AI is the same. We have seen remarkable progress, but too many are treading the same path of a single-use; a single fixed-function application. This may deliver some outcomes, but it isn't designed for flexibility, speed, and extensibility.
- It doesn't ensure that users are equipped to tackle today's challenges, while preparing for tomorrow's needs. First adopters may initially go for fixed-function solutions, but these are not suitable for mature adoption that requires scalability, customizability, security, and extensibility, to future-proof and deliver greater ROI.

### AI Vision: The 'next big thing' will soon be supplanted

LLMs dominate AGI conversations, but big things are taking shape in vision. Text is easier as it is structured by semantic and grammatical rules. Vision data is more complex and heavier, with cumbersome workloads and bigger files. Edge computing is required for vision at the source: this is why vision is the next big thing.

Sight is crucial to human intelligence. Teaching machines to see unlocks a deeper and more intuitive understanding of the world. We seek to see the world differently through AI Vision, through the perspective of our clients and the challenges they face tomorrow.

We are pioneers of game-changing technology that places organizational needs at the very center of our thinking. An intelligence engine that drives positive outcomes through data. A paradigm where cameras don't just detect objects but rather truly understand what they see and what to do about it. An approach that adapts to a wide range of use cases with minimal model training and the ability to embed AI models into real-world applications to solve business problems.

While everyone talks about AGI, we have built VI. We have paved the way for VGI, AGI's cousin, specializing in visual tasks.

We are equipping machines with the ability to see as clearly as we can, if not better, solving every vision-related challenge, and reshaping entire industries in the process.

## (B) Turning point: From CV to true Visual Intelligence

We are in a new era of computer vision: true VI. Powered by deep learning, emergent foundation models, and the wide availability of affordable, high-performance hardware, VI is no longer confined to research labs or highly controlled production lines.

Businesses across industries stand at the threshold of deploying VI at scale to solve real-world problems: from quality inspection in manufacturing to safety enforcement on construction sites. Yet not all computer vision technologies are created equal.

Traditional vision systems - reliant on rigid pixel comparisons and costly data pipelines - struggle to adapt beyond narrow, predefined tasks. Even those more flexible, deep learning solutions depend on vast amounts of annotated data and long training cycles.

This is where a transformative concept is now starting to take center stage: the emergence from VI of VGI.

### Beyond static models: The triumphant emergence of VGI

VGI is a next-generation framework for computer vision: one that moves beyond static, task-specific models. Instead of training a new model for every use case, VGI enables AI systems to adapt dynamically to new visual tasks, just as humans do.

Think of it as giving machines the ability to “see” and interpret the physical world with flexibility and context-awareness. If Artificial AGI aims to generalize intelligence across domains, VGI focuses on the visual realm.

VGI empowers AI to handle any vision-based problem, even under changing conditions or with minimal retraining. VGI systems adjust to different camera angles, lighting environments, or edge cases. This redefines what’s possible for real-world AI vision.

## (C) DeepSeek's leapfrog: A case study in AI cost engineering

### What did DeepSeek achieve and how did it do so?

DeepSeek's breakthrough represents a paradigm shift in AI development economics that mirrors the disruptive potential for visual intelligence systems.

The startup achieved ChatGPT-level performance with only $5.6m in **development costs** to over $3bn for GPT-4. At the same time, DeepSeek **outperformed ChatGPT** in specific domains like mathematics, achieving 90% accuracy versus ChatGPT's 83%.

The company accomplished this through innovative architectural choices, including '**mixed precision**' - full-precision 32-bit floating point numbers (FP32) with low-precision 8-bit numbers (FP8) and 671B-parameter '**Mixture-of-Experts**' (MoE) architecture that was trained for 1/18th the cost of GPT-4o.

The cost advantages extend to **deployment**, DeepSeek-R1 priced at $0.55 per million input and $2.19 per million output tokens, compared to OpenAI's GPT-4o at $15.00 and $60.00 (as well as 90% less energy consumption and 92% lower environmental impact).

### The implications of DeepSeek for CV and Visual Intelligence

This breakthrough suggests that similar cost-efficiency revolutions could be imminent in VI. DeepSeek's approach demonstrates how specialized architectures, efficient training methodologies, and open-source strategies can dramatically reduce the computational and financial barriers to advanced AI capabilities.

For computer vision, this would mean that sophisticated visual intelligence systems - currently requiring massive GPU clusters and millions in training costs - could be developed and deployed at a fraction of current expenses.

The implications are profound:

- Start-ups could develop custom VI models for specific industries without prohibitive capital requirements
- Edge deployment of advanced vision systems could become even more economically viable, and
- The democratization of visual intelligence could accelerate innovation across manufacturing, healthcare, autonomous systems, and beyond

Just as DeepSeek challenged the assumption that frontier AI requires massive budgets, similar breakthroughs in VI architectures could make advanced computer vision capabilities accessible to a vastly broader range of applications and organizations, potentially accelerating the timeline toward VGI by orders of magnitude.

> While we're stuck debating whether machines can think like humans, we are missing the profound truth that the real revolution won't come from artificial minds that can reason... but instead from artificial eyes that can see better than humans ever could.
>
> We live in a world drowning in the most complex form of data - visual information - and we just cannot process fast enough: health and safety incidents and fatalities we could easily and simply prevent; supply chains we can't monitor; infrastructure we can't maintain; climate patterns we can't track; opportunities staring us all in the face.
>
> **Gaudenz Boesch,**
> Co-Founder and Co-CEO @ viso.ai

## (D) Modularity and building blocks: Technological enablers of VGI

Modularity is needed because of the technically diverse components needed to unlock VGI. This is the key building block in implementing a future-proof system design.

The emergence of VGI has been made possible by several converging technological breakthroughs that have fundamentally shifted the landscape of computer vision.

The advent of foundation models has demonstrated how large-scale pre-training can create adaptable systems that generalize across diverse tasks with minimal fine-tuning. Similarly, Large Vision Models (LVMs) have brought this foundation model approach to the visual domain, enabling VGI systems to tackle multiple vision tasks without the traditional requirement for extensive task-specific training.

These models serve as the intelligent backbone that allows machines to understand visual contexts with unprecedented flexibility. Equally critical has been the widespread availability of affordable, high-performance hardware that makes edge deployment economically viable. Modern GPU architectures and specialized AI chips have democratized access to the computational power necessary for sophisticated visual processing, while advances in edge computing have brought AI closer to data sources.

This hardware evolution, combined with innovative architectural approaches like mixed precision frameworks and Mixture-of-Experts (MoE) models, has dramatically reduced both the computational and financial barriers to deploying advanced vision systems. The convergence of these technological enablers has moved VGI from research laboratories into real-world applications across industries.

## (E) Ripples in the pond: The technological implications of VGI

The technological implications of VGI extend far beyond improved computer vision capabilities, representing a fundamental shift in how AI systems are architected and deployed. Unlike traditional vision systems that required separate models for each use case, VGI enables dynamic adaptation to new visual tasks through techniques like self-supervised and weakly supervised learning.

This reduces dependency on vast labelled datasets and eliminates the need for lengthy retraining cycles. This makes advanced vision capabilities accessible to organizations that previously lacked the resources for custom AI development. The integration of human-in-the-loop feedback mechanisms creates self-improving systems that continuously enhance their performance through real-world deployment.

The broader technological implications reshape the entire AI development paradigm, moving from static, task-specific models to adaptive, general-purpose systems.

VGI's end-to-end integration capabilities establish a new standard for AI system architecture that emphasizes flexibility and scalability, spanning data input, deployment, monitoring, and continuous improvement. This shift enables edge-based computing at unprecedented scales, with centralized management of distributed hardware allowing one-click deployments across thousands of devices.

The democratization of advanced vision capabilities through VGI platforms could accelerate innovation across manufacturing, healthcare, autonomous systems, and beyond. This could potentially compress the timeline toward VGI by orders of magnitude, all the while making sophisticated AI accessible to a vastly broader range of applications and organizations.

## (F) From pixel matching to adaptive understanding

Historically, computer vision relied on strict pixel-based rules. These performed well in environments with controlled lighting and standardized layouts, like product lines in a factory. Yet once any variable deviated from expectations, these models faltered. New shapes, lighting changes, or even dust could trigger false positives or system failures

## (G) Deep Learning: A step forward but far from the finish line

DL brought significant improvement. Convolutional Neural Networks (CNNs) and other architectures allowed for more robust pattern recognition. These systems generalize better across image variations. But their success came at a price: large, high-quality labelled datasets, GPU-heavy training cycles, and months of fine-tuning.

## (H) Don't look back: Why VGI is the No.1 game changer in AI

VGI represents a step change in how vision systems are developed, deployed, and improved. We see four dimensions to this:

### *1. Adaptability: Suitable for any use-case*

VGI models can adjust to new tasks without retraining from scratch.

### *2. Data-layer: Self-supervised learning*

Towards self-learning: 10x learning cycles, moving faster with less data .

### *3. Real-time feedback: Human-in-the-loop by design*

VGI integrates human-in-the-loop feedback, letting users improve performance directly from the application layer.

### *4. End-to-end integration: Closed stack, covering data input, deployment, monitoring, and improvement*

VGI isn't just a model: it's a complete stack, covering data input, deployment, monitoring, and improvement.

## (I) On terra firma: The five pillars of a VGI framework

To understand VGI's power, we must look beyond the model. Real-world vision systems require a suite of interconnected capabilities. Here are the five core components of a VGI architecture:

### *1. Model: Large Vision Models (LVMs) for next-gen visual understanding*

The AI model is responsible for interpreting visual input, be it detecting defects, identifying safety violations, or tracking movement. Thanks to large vision models (LVMs), many tasks can now be tackled with minimal fine-tuning. Yet the most capable model addresses only a fraction of what's needed to deploy a functioning system.

### *2. Application layer: Turning data into meaningful business impact*

Applications define how visual insights are used. For example, what triggers an alert, how video feeds are processed, and where does the output go? Without this layer, a model is just a proof of concept. VGI platforms provide application-building tools, allowing developers to define inputs, logic, and outputs with minimal coding.

### *3. Deployment architecture: Instant processing at the source*

Vision applications process massive data volumes - especially when hundreds of cameras are involved. Streaming every frame to the cloud can be prohibitively expensive and slow. VGI enables edge-based computing, bringing AI closer to the data source. Modern platforms offer centralized management of distributed hardware, enabling one-click deployments across thousands of edge devices.

### *4. Configuration tools: Scale across sites instantly*

Each deployment environment is unique. A construction site has different risks than a food processing plant. VGI platforms offer remote configuration for camera-specific settings - like defining zones of interest or setting detection thresholds. These settings are stored as digital twins, enabling rapid replication across devices and locations.

### *5. Continuous interaction (human-in-the-loop): Adaptive learning*

When AI misclassifies a defect or misses an incident, humans need to intervene. With VGI, user corrections are fed back into a central knowledge base: the system learns. This self-improving mechanism reduces error rates and broadens task coverage.

Active learning occurs by tagging incidents and adding comments, moving them across in the incident center, or flagging incorrect detections. The system then uses those data points as feedback. Truly intelligent vision systems build components where the user can define pain points that impact the insights from analyzing application output data.

## (J) Dashboards and analytics: Insights driving behavioral change

Beyond real-time alerts, organizations need tools to understand trends. VGI systems automatically aggregate event data into dashboards that visualize metrics like safety compliance, incident frequency, and quality KPIs. Teams can also capture annotated video snippets, useful for audits, insurance claims, or compliance reports. These features turn vision systems into strategic assets for business intelligence.

# 4. Welcome to the future: The dawn of VGI

> Artificial intelligence will reach human levels by around 2029. Follow that out further to, say 2045, we will have multiplied the intelligence, the human biological machine intelligence of our civilization a billion-fold.
>
> **Jürgen Schmidhuber,**
> Scientific director of the Dalle Molle Institute
> for Artificial Intelligence Research

# 4. Welcome to the future: The dawn of VGI

## In summary: Focus of this section

*This chapter looks ahead to the future of Visual Intelligence and its convergence with AI and broader technological advances. It charts how innovation across the 2020s–2050s - from multimodal grounding to embodied agents, from human-computer symbiosis to sentient AI - will reshape civilization. VGI sits at the center of this transformation, evolving from perception to prediction and ultimately to visual super intelligence. The result is an integrated future where vision, intelligence, and technology merge to augment human capability and redefine the very nature of work, safety, and progress.*

---

In order to step boldly into the future of Visual Intelligence, we believe it is essential to outline developments we anticipate both in the technology sector at large and across AI that will pave the way to VGI.

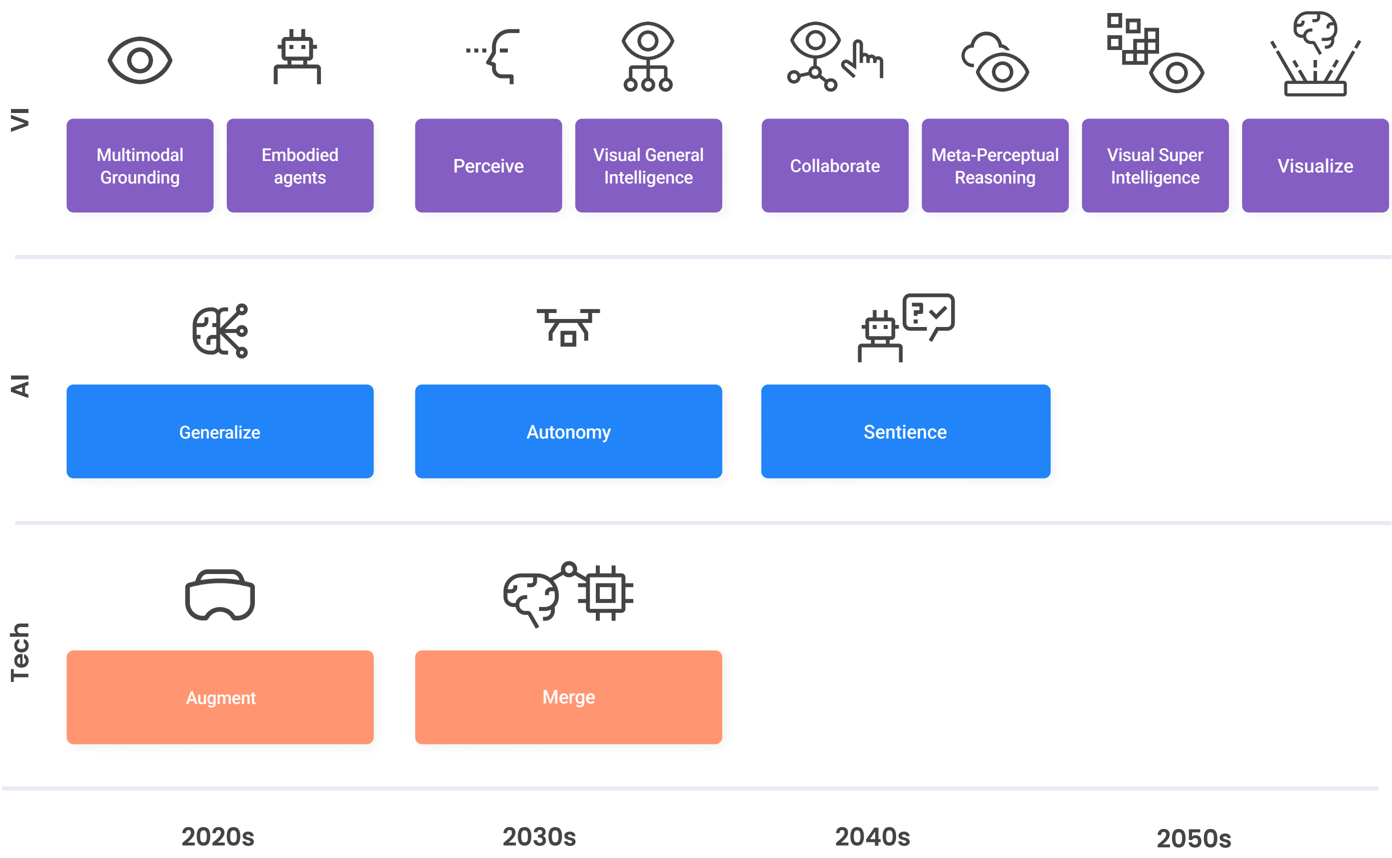
## The future of Visual Intelligence: Inside the crystal ball

This timeline illustrates the convergence of three technological domains - VI, AI, and general technology - across a period from now through to the 2050s. It shows how:

- Broader technology advances from augmentation to transcendence
- AI progresses from generalization to symbiosis, and
- Visual capabilities are evolving from multi-modal processing to advanced VI

This culminates in an integrated future where these domains merge to transform and reshape human civilization.

### The viso.ai roadmap to the 2050s



## The future of technology: Two decades of evolution

The table below outlines a two-decade technological evolution that progresses from augmentation to transcendence, each phase fundamentally reshaping human civilization, enabled by innovation and advancement.

From our current theme of "Augment", where AI, wearables, and IoT create a merged physical-digital reality that redefines how we interact with our environment, the 2030s shift to a "Merge" phase. In establishing human-computer symbiosis through neural interfaces and bioelectronics, the boundaries may become blurred between user and machine while enhancing cognitive capabilities.

### Table 1 - The future of technology: 2020s - 2030s

| Decade | Theme | Description | Why it matters |
|---|---|---|---|
| **2020s** | **Augment** | Merging of physical and digital with wearables, IoT, and AI overlays | Redefines interaction with the world through AI-assisted tools and interfaces |
| **2030s** | **Merge** | Human-computer symbiosis emerges via neural interfaces and bioelectronics | Redefines the boundary between user and machine, enhancing cognition and control |

## The future of AI: The shift from multimodal models to self-aware systems

The table charts a thirty-year progression of AI development from generalization to symbiosis, marking humanity's evolving relationship with AI systems.

The 2020s are characterized by the theme of "Generalize," where multi-modal AI and foundation models create flexible, context-aware capabilities, setting the stage for AGI.

The 2030s shift to "Autonomy," deploying general-purpose AI agents to manage homes, work, and physical tasks, expands human productivity via real-world AI integration.

And the 2040s enter the contentious era of "Sentience" where AI systems display introspection and self-awareness, forcing society to grapple with fundamental questions about AI consciousness, rights, and ethics.

This trajectory represents a transformation from humans being AI users to AI partners, ultimately reshaping the nature of intelligence and decision-making in our civilization.

### Table 2 - The future of AI: 2020s - 2040s

| Decade | Theme | | Description | Why it matters |
|---|---|---|---|---|
| **2020s** | **Generalize** | | Foundation models, multimodal AI, and agents emerge | This emergence is the path to generalized, flexible, context-aware AI capability |
| **2030s** | **Autonomy** | | General-purpose AI agents manage homes, work, and physical tasks | Expands human productivity and autonomy through real-world AI deployment |
| **2040s** | **Sentience** | | Debate intensifies over conscious or self-aware AI as systems show introspection | Forces redefinition of rights, ethics, and societal structure in relation to AI |

## The future of VI: A springboard to Visual Super Intelligence

The table below envisions a speculative yet clearly structured timeline outlining the advancement of computer vision from 2025 through the 2050s.

Each decade is associated with a visionary theme - from “Multimodal Grounding” and “Embodied Agents” to “Visual General Intelligence”, “Collaborate,” and ultimately “Visualize”, highlighting how visual systems are expected to evolve from basic perception to high-level simulation and cognitive visual synthesis.

From the near-term fusion of visual input with LLMs to long-term capabilities like brain-guided holography and hypothesis-driven perception, this forecast captures the anticipated shift from passive image analysis to dynamic, immersive, and human-like visual intelligence.

### Table 3 - The future of VI: 2020s - 2050s

| Decade | Theme | Description | Why it matters |
|---|---|---|---|
| 2020s | Multimodal Grounding | Vision integrated into Large Language Models for visual reasoning | This reasoning builds a bridge between perception and conceptual understanding |
| | Embodied Agents | Agents navigate environments with learned visual memory and active control | Unlocks enhanced robotic autonomy and spatial interaction between agents |
| 2030s | Perceive | Vision systems process and interpret environments with near-human accuracy | Enables full immersion in AR/VR and intuitive human-robot collaboration |
| | Visual General Intelligence | Systems perform open-world visual reasoning and decision-making | Forces redefinition of rights, ethics, and societal structure in relation to AI. cross-domain cognitive vision. |
| 2040s | Collaborate | Vision-guided agents operate seamlessly in unstructured physical worlds | Moves CV beyond passive observation to dynamic interaction in complex spaces |
| | Meta-Perceptual Reasoning | Systems understand what could or should be seen to guide perception | This understanding introduces hypothesis driven visual intelligence |
| 2050s | Visual Super Intelligence | VSI perceives and simulates with high-dimensional fidelity and abstract representation | Enables a new class of AI that sees, imagines, and creates beyond human vision |
| | Visualize | CV systems reconstruct, simulate and visualize from mental input and minimal data | Unlocks thought-driven design, creative augmentation, and neuro-visual feedback. |

# 5. The ‘North Star’: Proving VGI

> “The next frontier in computer vision is moving from perception to understanding and interaction.
>
> **Sanja Fidler,**
> Vice President of AI Research at NVIDIA

# 5. The 'North Star': Proving VGI

## **In summary:** Focus of this section

Finally, we introduce VGI as a system with general-purpose visual cognition that rivals and ultimately surpasses human visual capabilities. It shows traits of a truly intelligent visual system - goal-directed perception and autonomous understanding, plus simulation and continual learning - proposing a new benchmark: The "Visual Turing Test" (evaluating indistinguishability for human vs machine visual reasoning).

It could compare medically and cognitively accepted human visual standards vs machine capabilities. While machines exceed biological limits in isolated areas (frame rate or spectrum detection), the leap to VGI lies in achieving generalization, memory, abstraction, and context hallmarks of human visual cognition.

## The 'North Star': Approaching VGI at the speed of light

We believe VGI will exhibit broad perceptual abilities and perform high-level visual understanding tasks such as interpretation, abstraction, and contextual recognition across diverse environments. It could perceive and reason visual scenes in dynamic, unstructured settings and apply its visual knowledge to unfamiliar situations:

### 1. General Visual Knowledge: Scene understanding and reasoning

- Possesses an expansive understanding of the visual world, across domains (objects, actions, environments, and materials)
- Can interpret and reason about a wide range of visual tasks without retraining or custom tuning (such as object detection, anomaly recognition, intent estimation, and spatial reasoning)

### 2. Goal-Directed Perception: Expert task-execution

- Visual processing and attention are aligned with purposeful tasks
- Can prioritize relevant features and perspectives based on underlying goals (such as safety, interaction, and optimization)

### 3. Autonomous Visual Understanding: independent decision-making

- Capable of making self-directed visual inferences without external prompts
- Identifies, tracks, and reasons about salient entities/events in complex scenes

### 4. Context-Aware Perception: Navigation and human-machine interaction

- Understands scenes not just as static images, but as parts of dynamic contexts
- Perceives spatial relationships, environmental conditions, self-location (if embodied), and the roles of other agents

### 5. Visual Agency: Intelligent agents

- Interacts with the visual world through perception-action loops (such as robotic manipulation, drone navigation, and AR/VR feedback)
- Acquires new visual understanding from interaction and environmental change

### 6. Robust Visual Generalization: Explorative agents and systems

- Adapts perceptual strategies to novel scenarios (from industrial inspection to underwater exploration, for example)
- Performs well in low-data settings and under varying visual conditions (including lighting, occlusion, scale, and clutter)

### 7. High-level visual cognition: Reasoning and prediction

- Infers intent, causality, and latent factors from scenes (such as predicting what will happen next and understanding why something occurred visually)
- Can generate and interpret abstract visual representations (e.g., maps, diagrams and sketches)

### 8. Continual visual learning: Always-on and never forgets

- Continuously updates visual knowledge: new environments, tasks, and objects
- Retains and refines previously learned capabilities without ever forgetting

### 9. The Visual Turing Test: The true benchmark of VGI

Could there be a 'Visual Turing Test' to clearly confirm achieving VGI? While Turing's test focused on indistinguishability in language-based interaction, a 'Visual Turing Test' for VGI might evaluate a system's ability to interpret and respond to real-world video or imagery. Crucially, were it to respond with the same precision, intent, and abstraction as a human observer, it could be a proof point for achieving VGI.

Might a VGI system, shown a complex scene, generate responses (verbal or action-based) indistinguishable from a human? This in turn, raises further, questions:

Could a system be explicitly trained to appear visually intelligent without truly understanding what it sees? How would we know/could we prove either way?

- What are the implications of the above for the future of the global economy?
- If a human could learn from VGI's visual reasoning - seeing the world through its 'eyes' - could that human surpass even the machine's visual capacity

> " The companies, the nations, and the world leaders who understand the potential of visual intelligence... they are not just preparing for the next technological shift: they are positioning themselves to solve humanity's most profound and greatest challenges.
>
> Because when machines can finally see and understand our world the way we do, but faster, better, and without fatigue that will be when we stop managing crises and start preventing them. We will stop reacting to problems and start anticipating solutions. Artificial intelligence in that sense becomes embodied artificial wisdom.
>
> The question 'Will Visual General Intelligence reshape our world?' is missing the point: your choice is whether you lead that transformation, or if not, watch it happen to you.
>
> **Jeremy Michaels,**
> Strategic Content Writer @ viso.ai

# 6. Conclusion:
# The age of sight - reimagined

# 6. Conclusion: The age of sight, reimagined

We are not merely witnessing an evolution in AI - we are crossing a visual event horizon. VGI is not a byproduct of AGI - it may well be the defining proof of its arrival.

VGI grants machines a faculty that has shaped civilization more than any other: vision. But unlike human sight, constrained by biology and subject to fatigue, VGI will offer relentless clarity, infinite memory, and perception untethered from time and space.

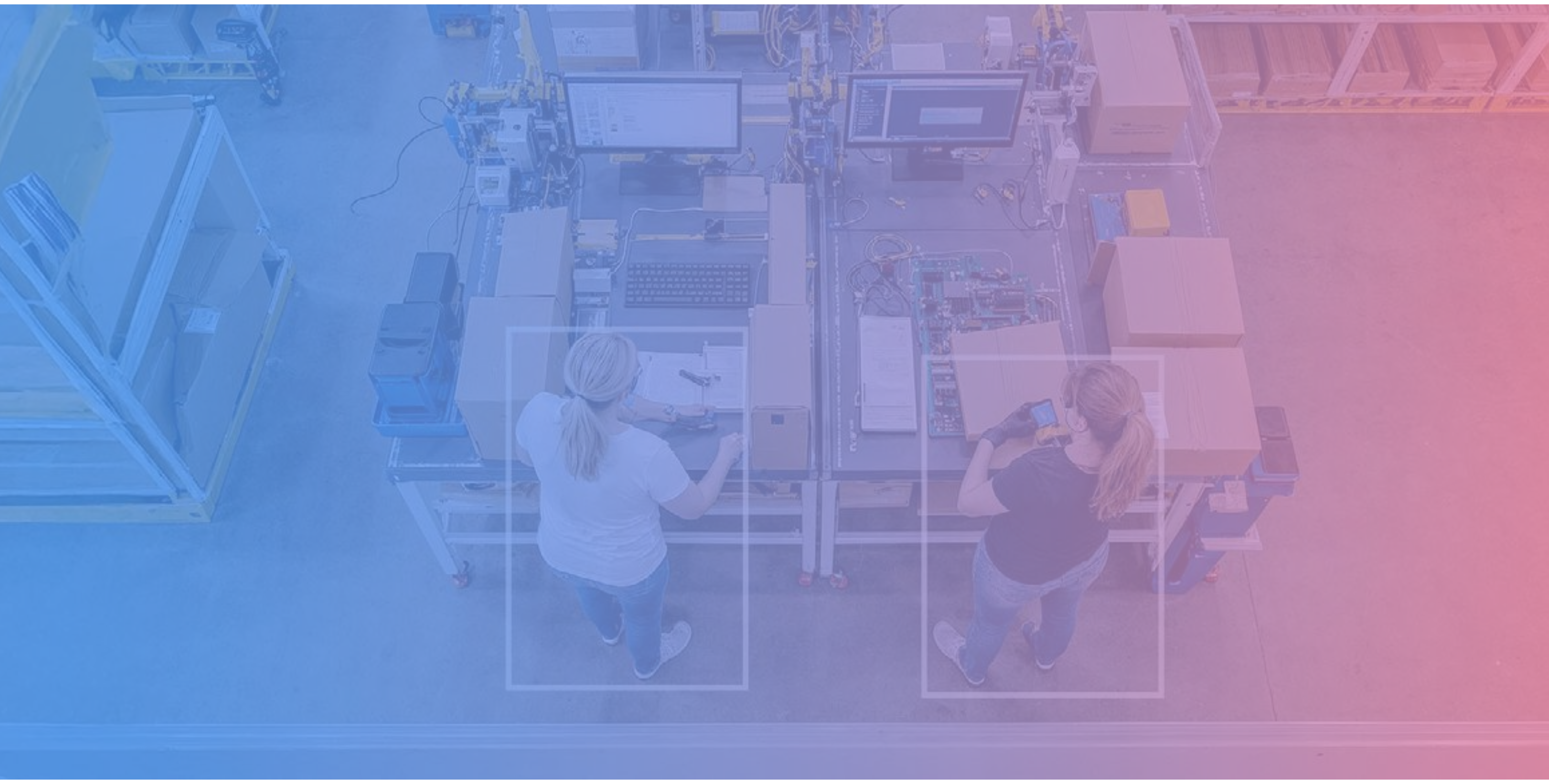
It will decode the visible and the invisible, understanding not only what is seen but what could be seen, redefining sight itself. This transformation is not just technical: it is civilizational.

With VGI embedded across industry, infrastructure, and everyday environments, every machine becomes a lens, every action becomes anticipatory, and every decision is informed by visual wisdom at scale.

That level of comprehension will not just eliminate inefficiencies and prevent disasters: it will rewire how we design, build, govern, and grow. To ignore the ascent of VGI is not just to fall behind. It is to remain blind in an increasingly sighted world. AGI may remain abstract, but VGI is a reality within reach, and its adoption is no longer optional.

If AGI promises minds that rival ours, VGI delivers the eyes that surpass them. The real revolution will not be heralded by sentient dialogue, but by a camera that understands.

Those who lead this visual awakening will not only shape the future of AI: they will shape the future through AI. And those who fail to act will not be disrupted. They will quite simply be left behind and rendered invisible.



  viso.ai

## Why viso?

**Viso Suite is the leading computer vision infrastructure for teams to build, deploy, and scale all use cases from prototype to production. Focus on tangible business outcomes, not infrastructure**

- **Build** powerful, custom AI vision applications
- **Deploy** real-time AI vision to the Edge or Cloud
- **Operate** and maintain a portfolio of applications
- **Scale** on robust infrastructure for any AI mode

---

## Get started

Contact us and speak with our team of experts to learn how you can get started with Viso Suite. Visit **www.viso.ai**


**Acknowledgments:**

*"The future belongs to those who believe in the beauty of their dreams"* (Eleanor Roosevelt)

This whitepaper is built upon the collective efforts of the team at viso, past and present. It is the culmination of many years of conceptual development, technological innovation and computer vision knowledge built up by key individuals.

The paper would not have been possible without research and insight provided in particular by the co-founders and co-CEOs of viso, Gaudenz Boesch and Nico Klingler. Several sections draw from thought leadership exclusively drafted by them.