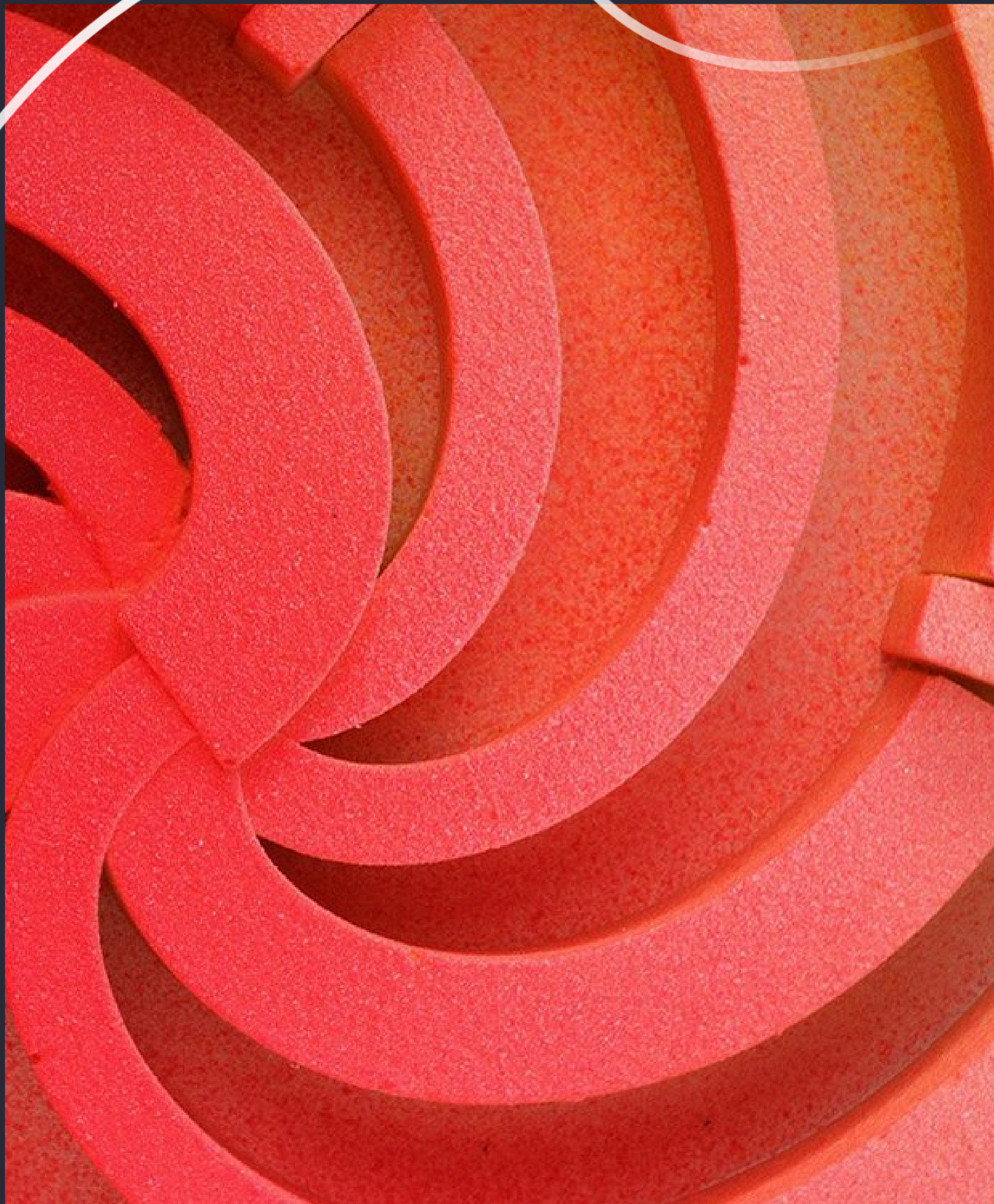
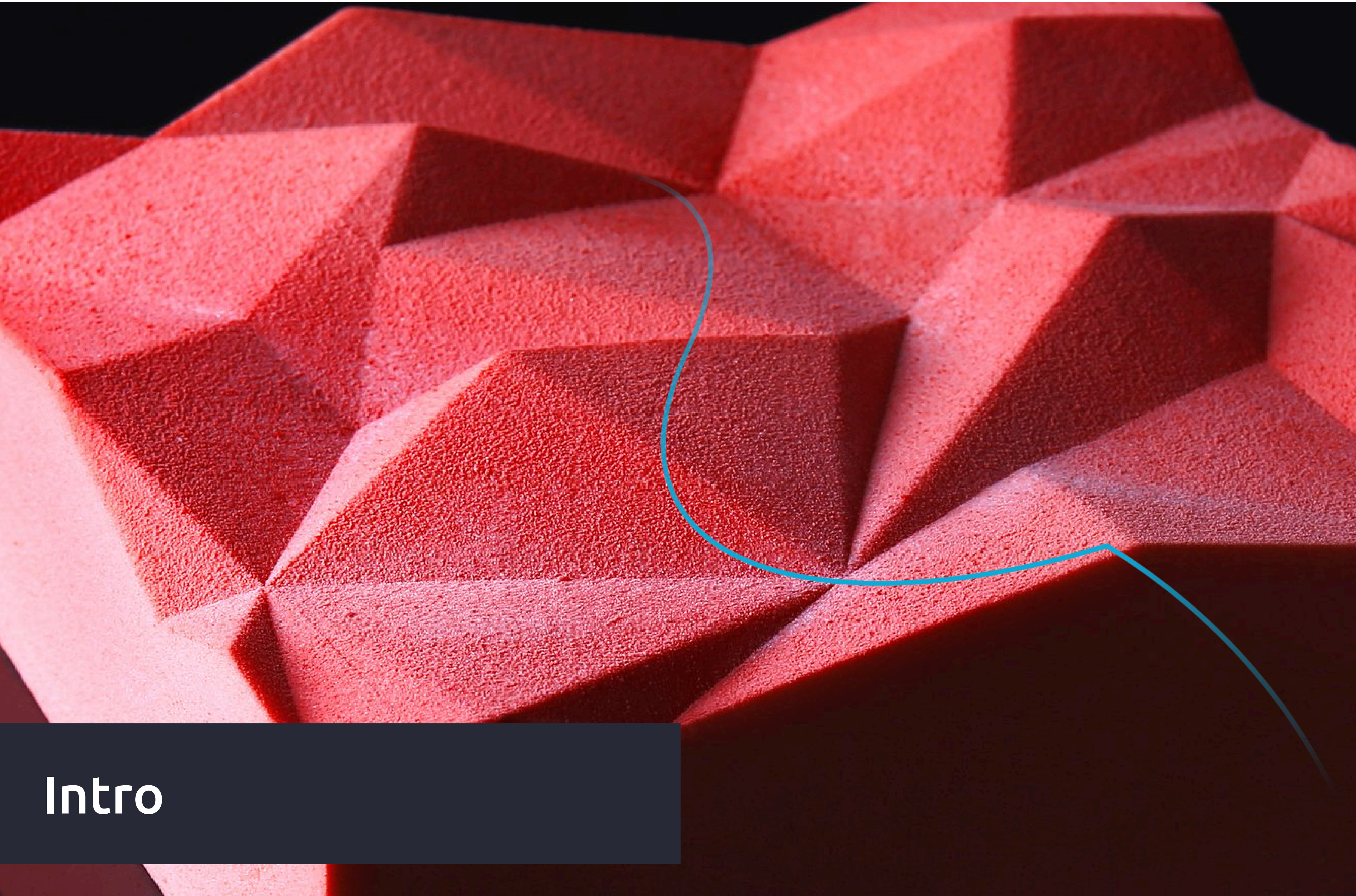
Gen AI is not enough

Why Gen AI must become Hybrid AI to augment engineering

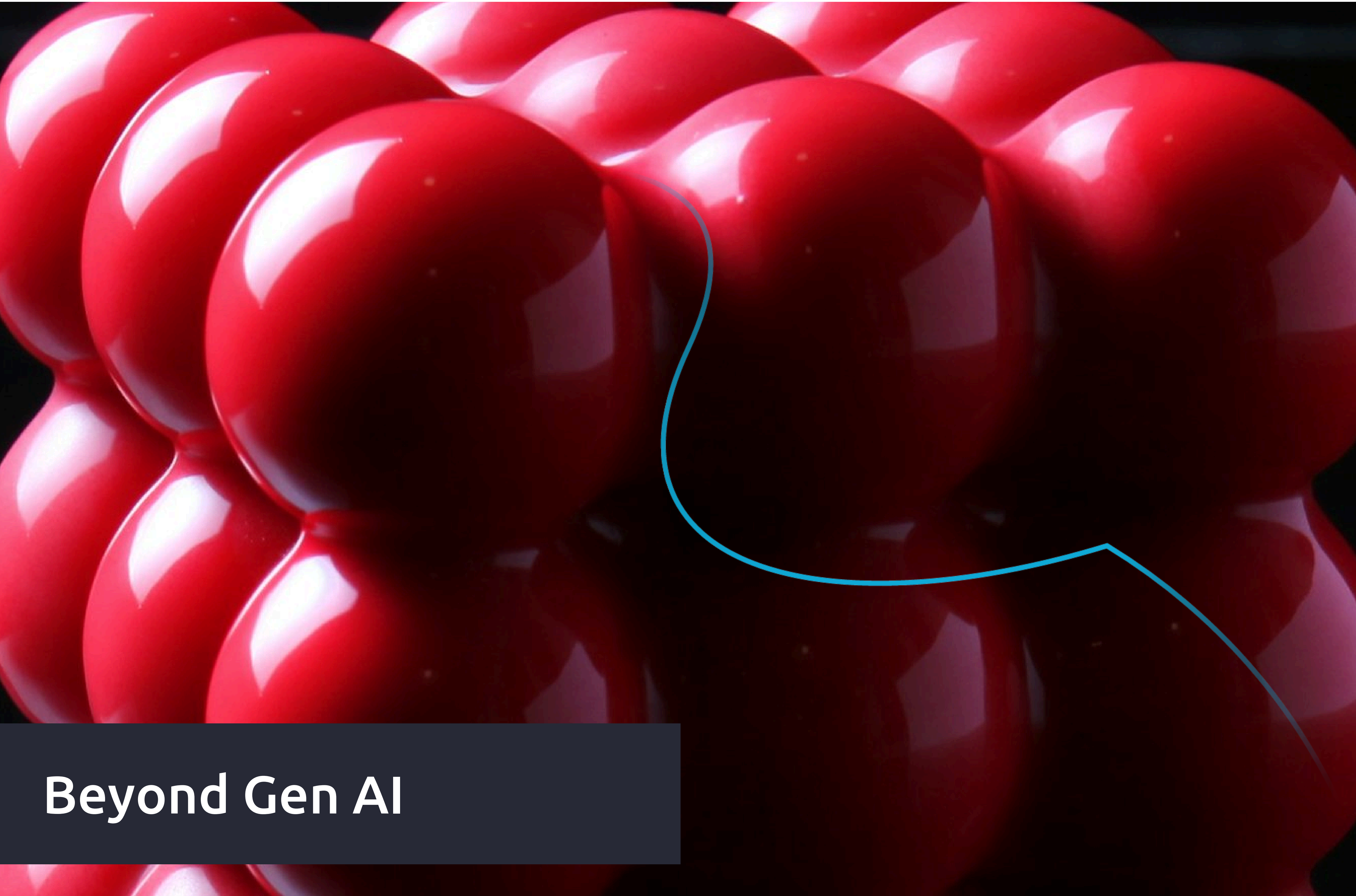




Intro

Over recent years artificial intelligence has been developing rapidly, and if anything, the pace is now accelerating. In generative AI (Gen AI) in particular, we're seeing ever-larger models and richer modes, encompassing text, image, video, speech, and code. We're seeing the evolution of multiple application areas, and an absolute proliferation of potential use cases. In most areas of business, the excitement is almost palpable.

Engineering is one such area – but it is a sector in which special rules and constraints apply. While creativity is essential, engineers face unique requirements: precision, regulation, industrial standards, and their enterprise's low risk tolerance – after all, this is a discipline that includes autonomous vehicles. Engineering applications often function 'in the real world' – where operating environments can be harsh (or disconnected from ecosystems) and computation requirements can be more limited than in cloud or on-premise implementations.



Beyond Gen AI

Engineering leaders need more than Gen AI alone can offer. They are inherently focused on solving real-world problems. They want to harness the power of AI to transform processes and also to create game-changing smart new products and services. But it needs to be AI that is fit for purpose – AI that can meet their demands for accuracy, specificity, quality, compliance, and dependability in *product,* not just desktop environments. As we'll see, Gen AI in isolation may help to achieve some early results, but it simply can't always satisfy these rigorous criteria. Thinking that does not deal with the specific failure modes of generative technology, will likely miss key engineering requirements.

Engineering: Why Gen AI alone is not enough

Engineering requires both precision and the ability to capture detailed data in many forms. There is a vast difference in engineering application between a photo-realistic video and the schematic diagrams of an airliner avionics system, even if both are represented graphically.

The requirement extends beyond text, image and video. Engineering organizations need:

- To factor in engineering diagrams, architecture, and process flows into the corpus of inputs and outputs of AI models
- To control precision, and the correctness of the system
- To be able to gauge the confidence levels in what AI is telling them
- To balance the innovation of AI against engineering's heightened requirements for things like safety, regulation, intellectual property, privacy, and security
- To act fast – but also, to build strategic foundations over the long term
- To embed AI not just in processes, but in trailblazing new products and services
- To achieve balance between the ability of AI to transform against sustainability commitments. For instance, ever larger language models require ever larger compute facilities to train them, which have significant energy requirements

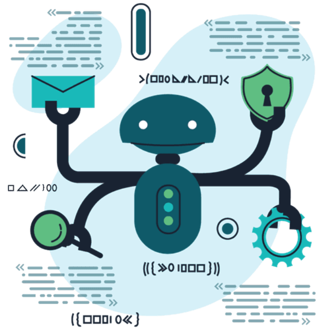
This is why Capgemini Engineering's approach combines generative machine learning models, which characterize patterns in data, with sources of structured knowledge and symbolic reasoning, including traditional symbolic AI, knowledge management, reasoning, and planning. We call it Hybrid AI.

It's an approach that means the value of AI is delivered with the quality and correctness the sector needs. Implemented on an industrial scale, it will change the way we approach mainstream engineering. It won't replace human endeavor, but rather enhance and accelerate it. It will, in other words, become Augmented Engineering – taking people, processes, products and services to new levels of intelligence and efficacy, while retaining both rigor and creativity.

What is *Hybrid AI?*

STATISTICAL MACHINE LEARNING

Current large machine learning models based on learning patterns in data, especially Generative AI are statistically impressive but individually unreliable.

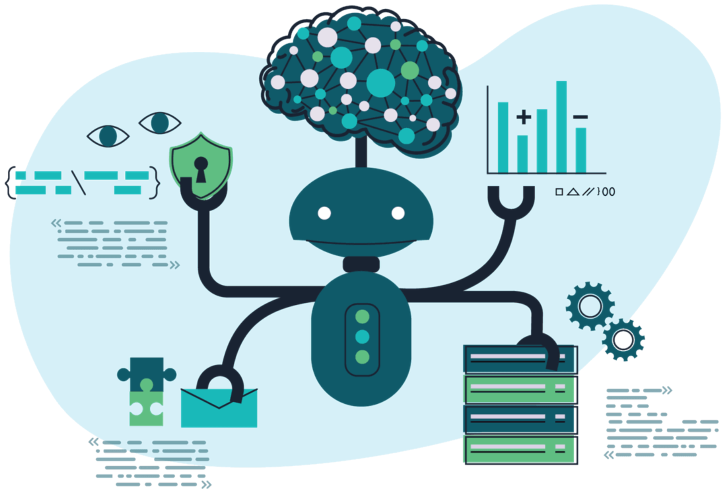


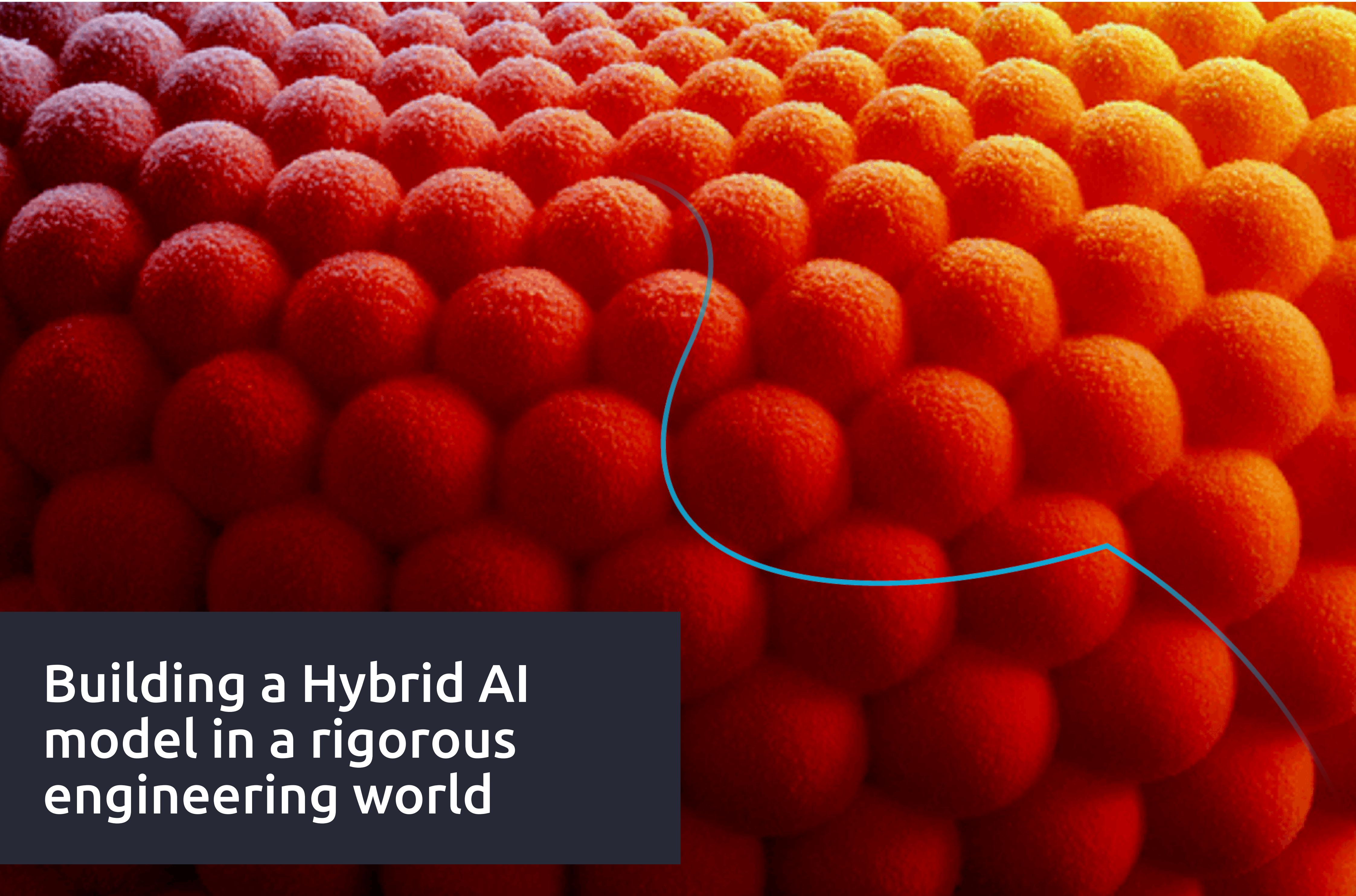
STRUCTURED KNOWLEDGE & SYMBOLIC REASONNING

Long history of more traditional symbolic AI knowledge management, reasoning, planning, logic.

HYBRID AI

Hybrid AI combines the best aspects of different types of AI and non-AI technologies. It can achieve outcomes that combine the value of AI with the quality and correctness required in engineering and science.





Building a Hybrid AI model in a rigorous engineering world

We all know that AI has the potential to make significant improvements to many industries, by making existing processes and products more efficient, as well as enabling whole new classes of applications.

However, in the world of engineering and science, those benefits are harder to realize, and for several reasons. First, because operating in this industry requires high degrees of precision, validation and traceability – traits that naïve AI is not known for. Secondly, engineering data is diverse and interconnected – not just text, but diagrams, CAD files, molecular descriptors and many more, and our solutions must be consistent across these modalities. Finally, engineers work in a diverse architectural ecosystem, physically deploying solutions onto many different sizes and scales of platform, often with challenging real-world constraints.

Hybrid AI is a new architectural weapon in the Augmented Engineer's arsenal which helps us unleash the power of AI into the engineers' world, without compromising on their core beliefs and convictions.

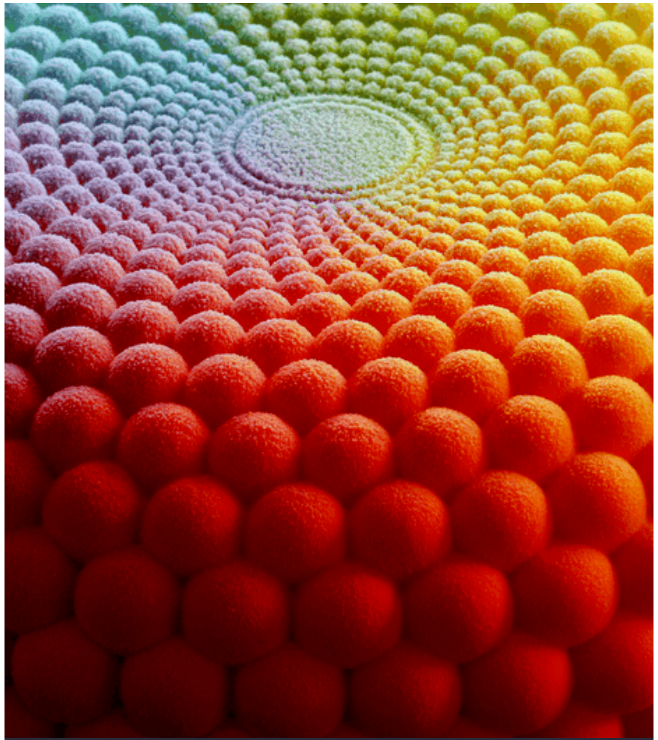
A Hybrid AI system combines AI models that derive their power from the powerful but unreliable statistical modeling of large data sets, with more traditional approaches to reasoning about structured knowledge, to produce hybrid systems that blend the best of both worlds. For example, an AI system for driving an autonomous vehicle needs to be able to do more than just spot road signs and other vehicles – it needs to be able to reason about the physics of how objects move or how human behavior changes in different cultures. Similarly, a generative design system for making aerospace parts will fail if it does not also consider the physical and structural properties of that component.

The practical reality of implementing these multi-faceted Hybrid AI systems is itself a multi-faceted challenge. Teams with expertise in contemporary machine learning methods need to come together with teams experienced in more traditional techniques for representing and reasoning about structured knowledge. It should be no surprise that both sides of this dichotomy are entirely reliant on data. While the Big Data era made claims that “data is the new oil”, the AI era has made that claim a reality, where those in possession of enough high-quality, well-curated data are able to convert that into a new tool with inherent value – an AI model.

In engineering, the data landscape is inherently multi-modal, where data is represented in different forms, for example, a written description of a process and a diagram of that process, and it is vital that consistency between those modalities is maintained. While for many, large language models have become a default paradigm, in this world, that narrow single-modality approach could be a limiting constraint.

Once the data is ready, training or fine-tuning a model can commence, but careful consideration needs to be given to the choices made here. Many different types of model are available, which have radically different capabilities, operational profiles and resource requirements. Choosing the wrong model could lead to massive operational costs, poor accuracy, or skeptical users. Larger models can also have significant environmental impact.

Hybrid AI again comes to the rescue here, allowing us to produce solutions that are much leaner. Why? Because the knowledge-driven part of the model can provide concise rules for how to behave when the pattern-based part of the model doesn’t have any similar prior experience. Purely pattern-driven models need to “see everything” in order to guarantee good general performance, a requirement which drives models to be larger and larger for diminishing returns. A hybrid model can provide sensible defaults for many cases, vastly reducing the need for large statistical models.



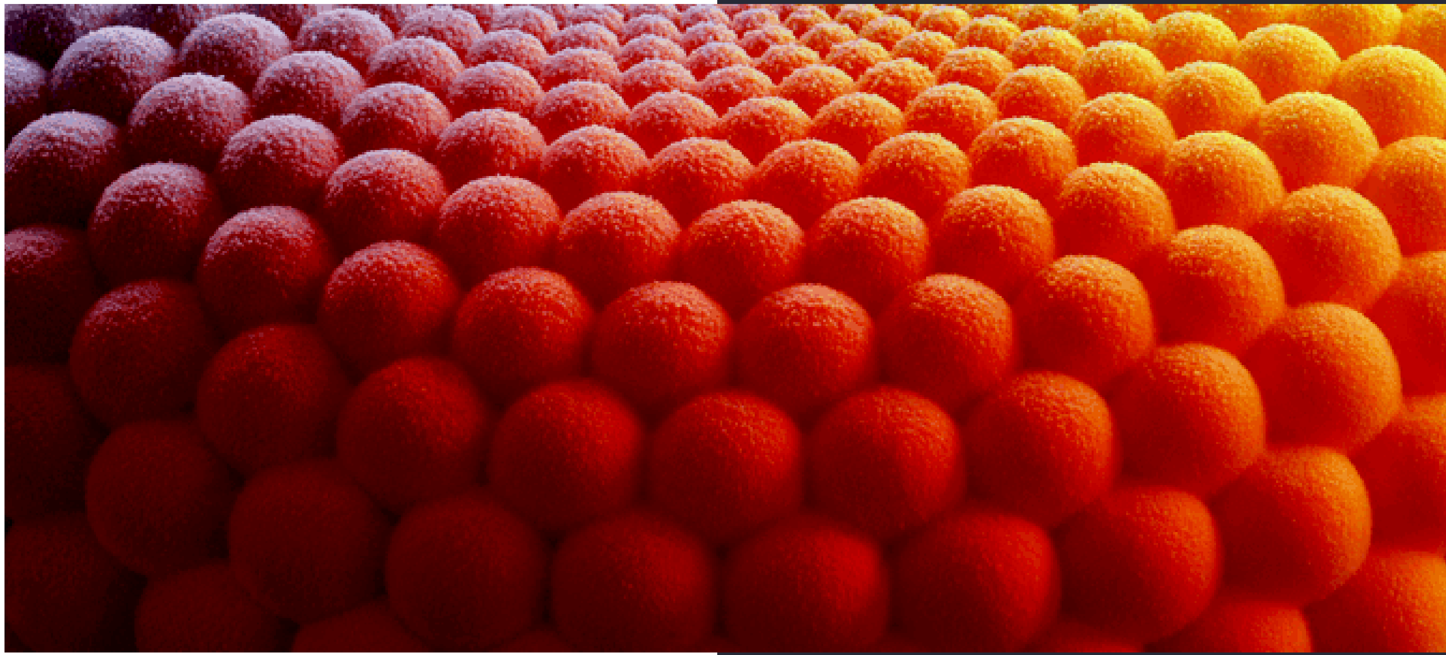
“Hybrid AI brings together generative machine learning models with sources of structured knowledge and symbolic reasoning. It is tailored to the world of engineering, where the data landscape is inherently multi-modal.”

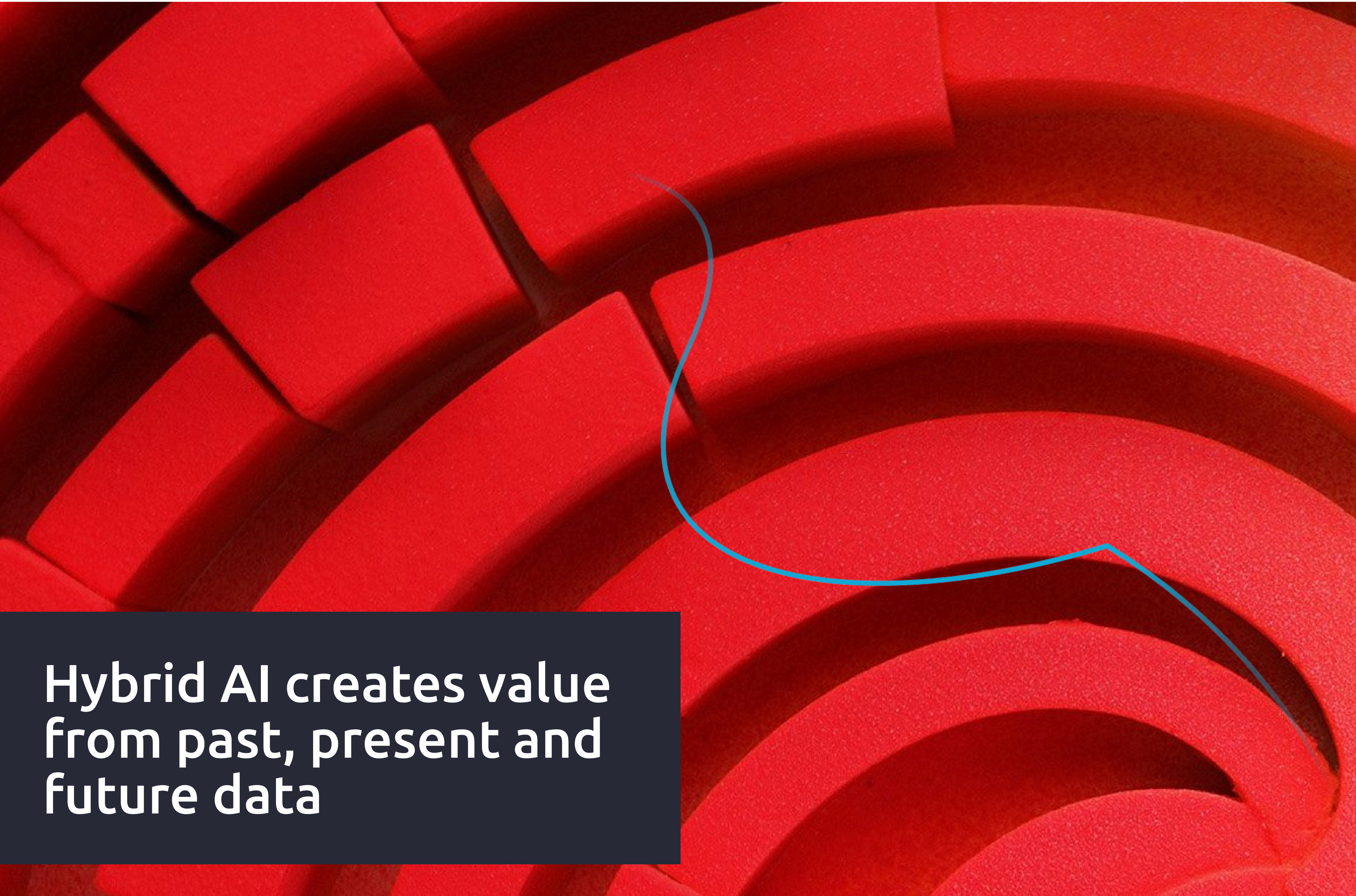
The Hybrid AI approach also helps to address one of the most insidious issues in AI deployments – that AI solutions (particularly Gen AI) fail in different ways to humans, which makes them unpredictable and untrustworthy. The harsh reality is that systems operating in the real world are judged not by how often and how well they succeed, but by how often and how badly they fail. These AI-specific failure modes must be managed, and engineering leaders need to decide on the scientific or engineering models or techniques they can use to control them and boost value.

To a large extent, choosing where to invest and determining what will be acceptable will depend on balancing a number of factors including through-life cost/benefit, competitive separation, what the supply chain is doing, what regulations prescribe, and what markets and society expect. For example, a one-off mechanical design solution that is highly carbon-expensive because of its computation, analysis and data costs may well be a cost worth paying if it reduces the weight of a commercial airliner fleet by 1%.

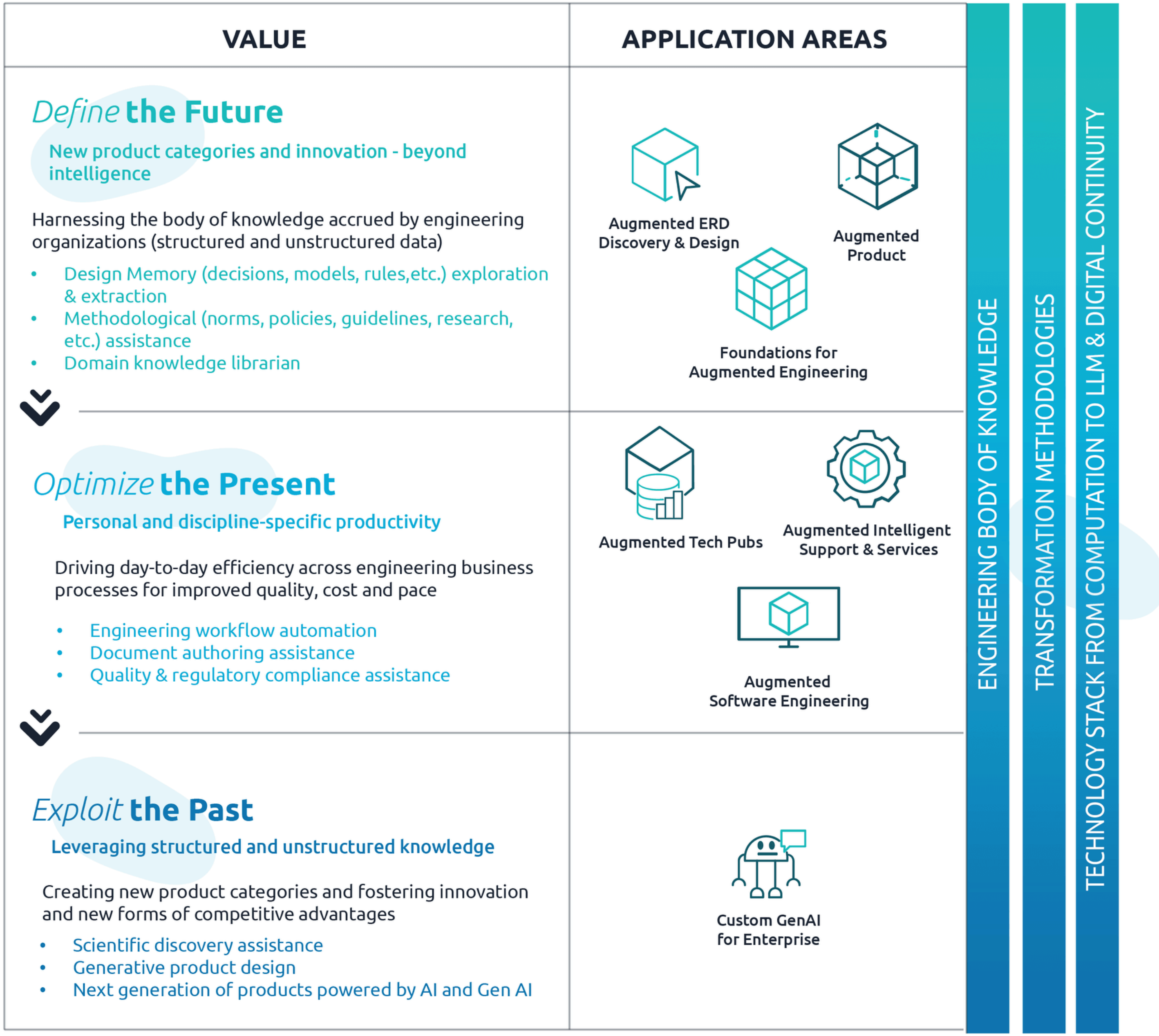
What ought to emerge from all these considerations is a combined model that is fit for engineering purposes – a model that draws not just on Gen AI, but on other disciplines, other types of AI, and sources of structured knowledge, to create a problem-solving instrument appropriate to the task at hand, that is transparent, trustworthy and reliable, and that delivers truly Augmented Engineering.

"Augmented Engineering will change the way we approach mainstream engineering. It will take people and processes, products and services to new levels of intelligence and efficacy."





Hybrid AI creates value from past, present and future data



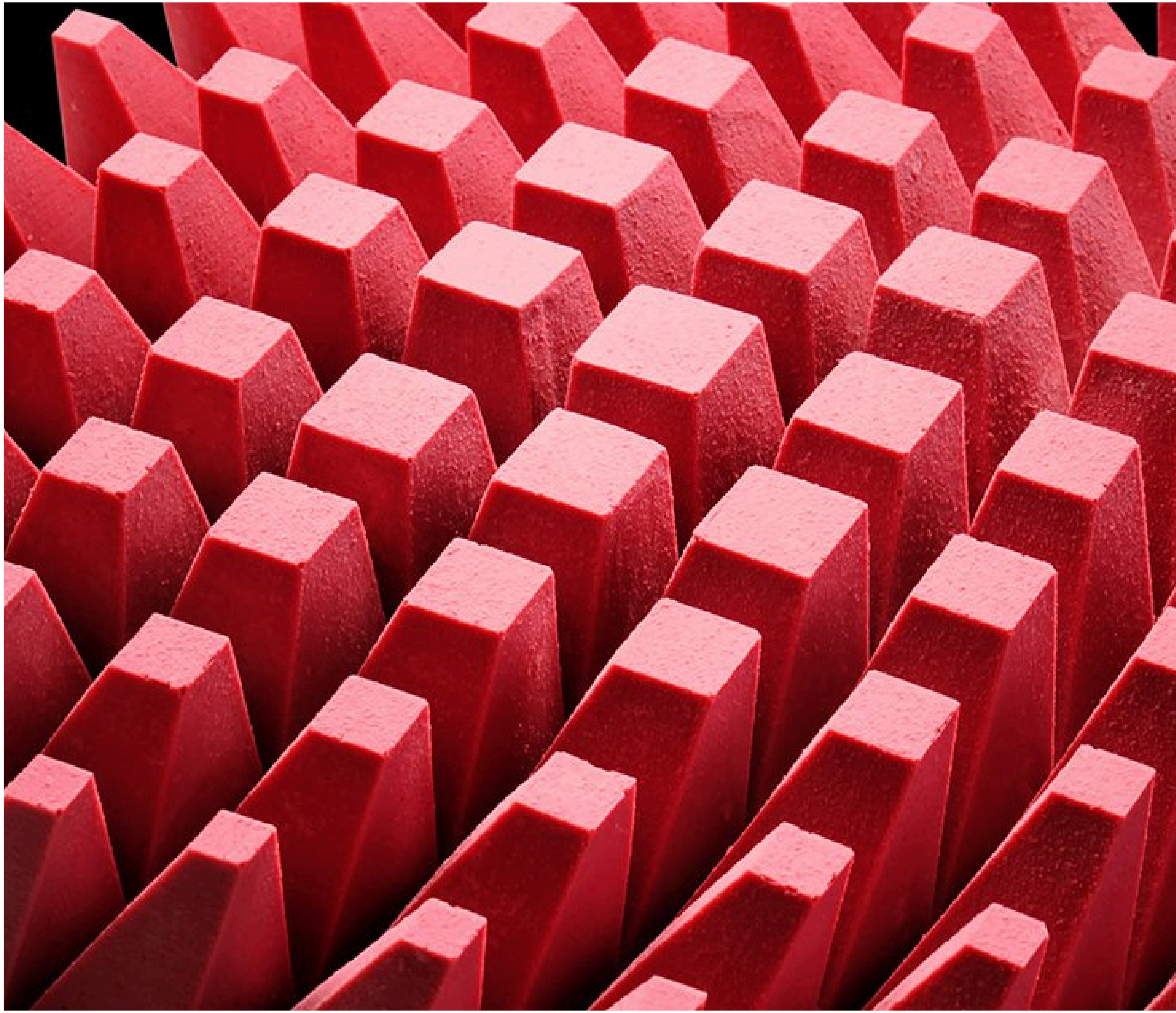
In deploying Hybrid AI, we should consider that its value and applicability extend though the life of a product or organization. Consequently, issues like scale, performance and data governance must consider the full data timeline.

Applying AI in engineering...

Much of the excitement around Gen AI has focused on the possibility of early results – on its ability, for instance, to write an executive summary of large amounts of information.

However, AI's capabilities shouldn't be considered as a series of isolated decisions or small-scale automation projects to address local process inefficiencies, but should rather be seen as an opportunity to step back, holistically assess the value that the business process should create, and seek a global improvement, by redefining activities from the ground up.

Engineers have this kind of pragmatism. It's a question of attitude – of the ability to think strategically as well as tactically (see 'Capgemini Engineering's mindset' at the end of this section). That's why we recommend that, in addition to being happy to take an early result, they should also be looking further ahead. In our view, it makes sense for engineering organizations to look for ways to create value over different time horizons for different levels of transformation.



... in the short term...

Let's look first at generating short-term value on the bottom line. This is where the ability of Gen AI to digest and synthesize complex information comes in. Combined with Retrieval-Augmented Generation (RAG), it enables engineers to use simple natural language routines to search and query knowledge bases or technical documentation. In manufacturing, this same Gen AI ability can accelerate classification and analysis of quality incidents. It can also accelerate generation of quality documentation in domains with heavy compliance requirements (such as life sciences) and automate the documentation of non-conformities. You can find other use cases here.

... in the mid term...

In the mid term, artificial intelligence can generate broader, deeper value on the bottom line. This is the domain of Hybrid AI, which brings Gen AI together with symbolic AI, reasoning engines, logic routines and more, and may take benefits beyond the point-tasks we have just seen, to broader classes of engineering – focusing, for instance, on augmenting the end-to-end scientific discovery process, rather than on summarizing research papers. It's also at this point that we can see intelligence becoming an integral part of products and services across organizations – from the cloud to the data center, from the office or factory floor to the edge.

... and in the longer term

In the longer term, Hybrid AI starts to change the nature of engineering in ways we may not yet be able to imagine, and in doing so, begins to deliver significant value to the top line. For example, it may absorb knowledge of industry tooling, processes, and norms from across the engineering organization, and use that knowledge to suggest entirely new products and processes, and new routes to achieving them cost-effectively and sustainably. It's at this point that Hybrid AI moves beyond 'doing the same thing, but better', and steps into exciting uncharted territory.

Our engineering mindset

Capgemini Engineering's mindset reflects the lessons learned from its extensive experience of running engineering projects. It's a mindset that aims to be time-responsive and context-sensitive, ensuring it remains relevant in an ever-changing technological landscape – and particularly in relation to the role of Hybrid AI in Augmented Engineering.

The mindset addresses several key areas:

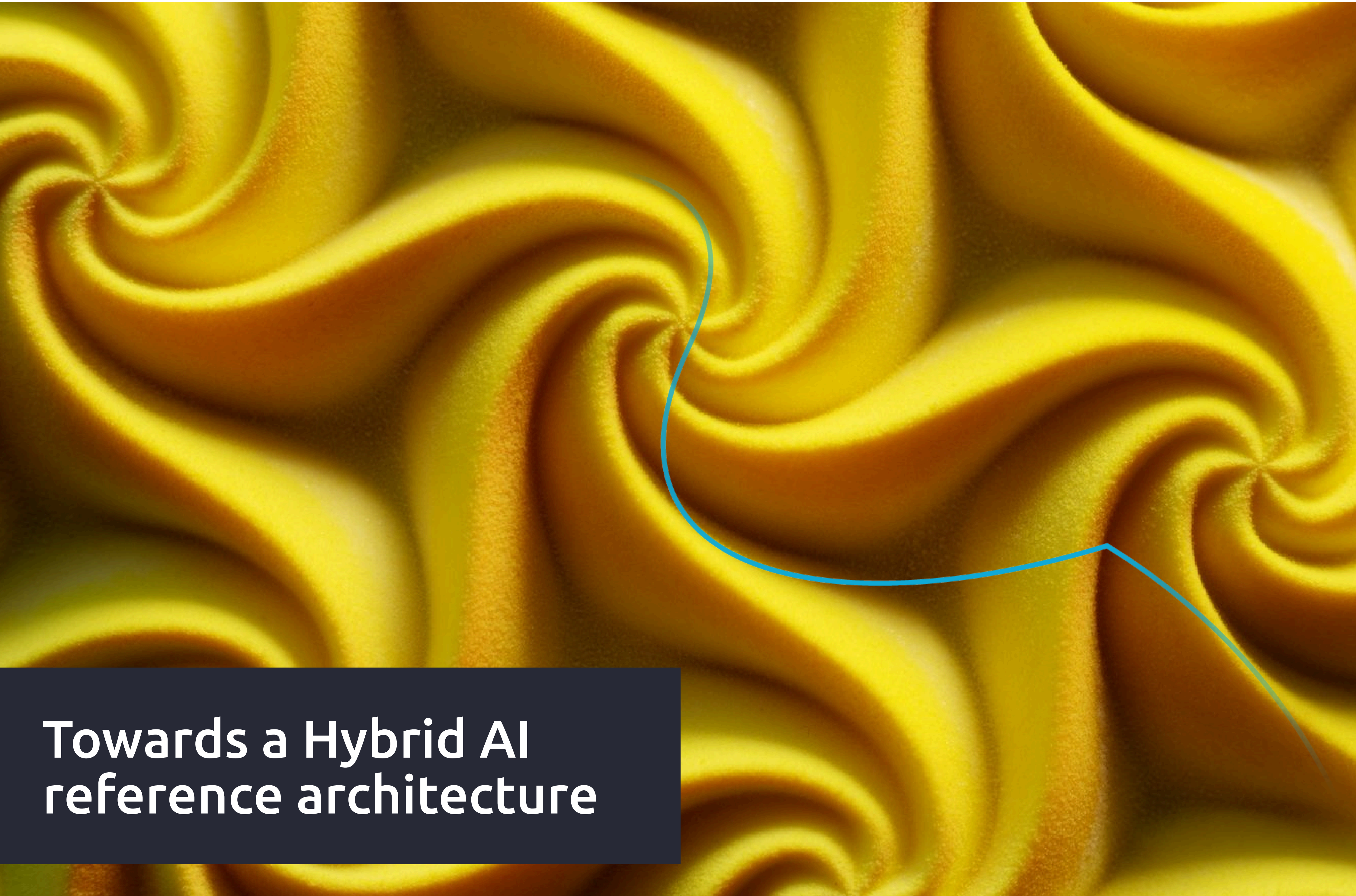
1. Artificial intelligence is transformational – a piecemeal implementation of AI, including Hybrid AI, will not succeed. We should seek comprehensive adoption of AI across the organization, focusing on transformational applications rather than isolated use cases

2. Treat sustainability as a systems issue – prioritize according to reliable data; integrate sustainability into every aspect of engineering practices to ensure long-term viability and environmental responsibility. Also, rational and responsible decisions about environmental impact must be made at the system level. For example, the impact of training and using a Gen AI model may be a tiny fraction of the benefits gained from reducing weight or energy consumption across the life of an aircraft or a truck

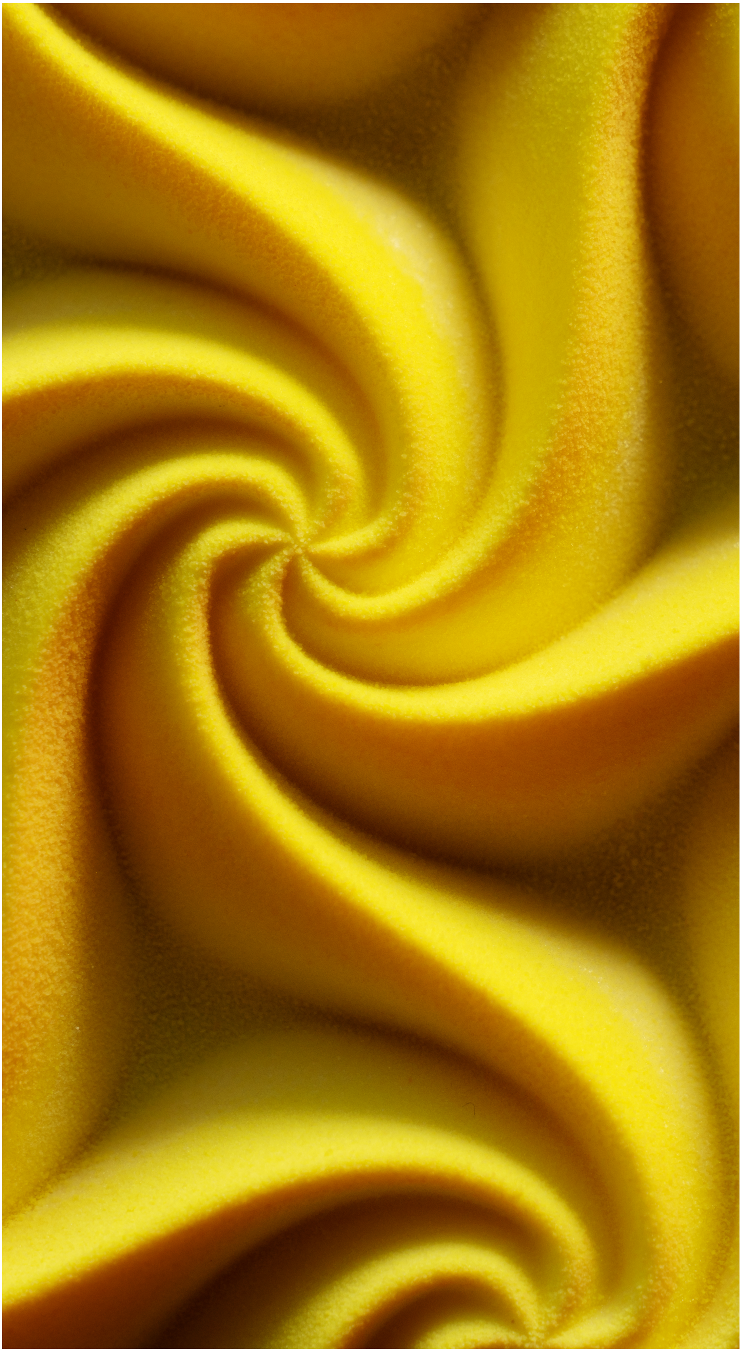
3. Agility through model-based engineering – exploit collaborative data models and mature tools to accelerate delivery and acceptance across disciplines, enhancing responsiveness and adaptability in project execution

4. Prioritize the role of the human in engineering – AI will transform products, but both its ultimate users and creators are human. Creativity and problem solving in development should be incentivized, while also placing human-centered design and social acceptability at the forefront of product development

5. Engage with key ecosystems – the transformative benefits of AI will cross organizational boundaries, contributing to the standards and tools that enable network effects



Towards a Hybrid AI reference architecture



It's a whole new world, setting new challenges of definition, integration, and coordination of technologies and activities, and that's why at Capgemini Engineering we believe a reference architecture is needed. Such an architecture should describe a standard approach for developing AI solutions, capturing a common understanding of key AI, Gen AI, and Hybrid AI concepts in an engineering context.

AI will shape and integrate with all aspects of engineering – the products that we build and the processes and facilities that we use to create them. To ensure that we develop consistent and efficient solutions, in our products, our design offices and our factories, we need consistent language, consistent structure and clear vision of the interfaces and roles of components – provided by a reference architecture.

Engineered products and engineering applications face constraints and expectations which make the AI landscape of engineering complex and constrained. A reference architecture must address functionality, connectivity, and data distributed across a range of physical and virtual environments – from hyperscaler data centers through network operator and enterprise facilities, to individual sites, equipment and portable devices. These will have different computational and physical constraints: of computation, storage, connectivity (bandwidth and latency), size, weight and electrical power.

Physical Model

The architecture of an AI application must resolve conflicting pressuress

CENTRALISED CLOUD	REGIONAL CLOUD/EDGE	ON-PREMISES EDGE	HEAVY ONBOARD EDGE	LIGHT ONBOARD EDGE
Multi-cloud interoperation	Cross-domain security	Deployable network core	Mesh networks	On Device

~100ms	LATENCY	<1ms
ZB	STORAGE	GB
1000t	WEIGHT	Kg
GW	POWER	mW
Exascale	COMPUTE	Gigascale

Foundation models | Private models | Private adaptors/RAG | Embedded AI | User/Device Interfaces

Pressure towards cloud

- Compute or storage intensive, more necessary for larger transformer models
- Data sharing & aggregation for large scale personalization
- Smaller & cheaper devices less able to support large models
- Load balancing, distribution & model management easier with central infrastructure

Pressures toward device

- Real-time applications requiring low latency, perhaps at the expense of AI performance
- Results required even if connectivity unavailable, unreliable, or slow
- Enhanced privacy (sensitive data does not leave the device)

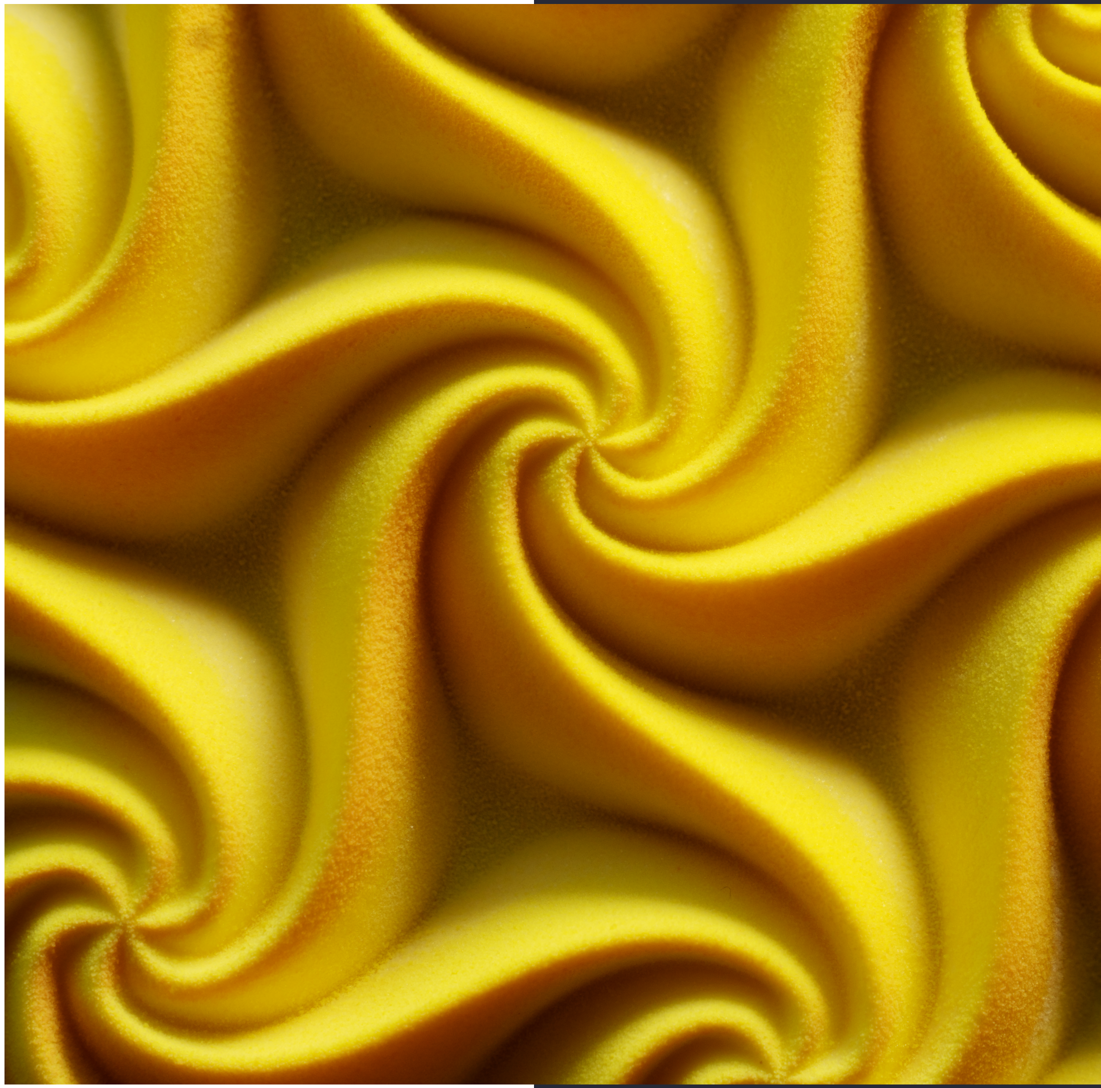
The Hybrid AI realizations will be similarly varied: from interfaces and lightweight inference in embedded devices, through federated learning and private knowledge bases to foundation models and their associated training infrastructure.

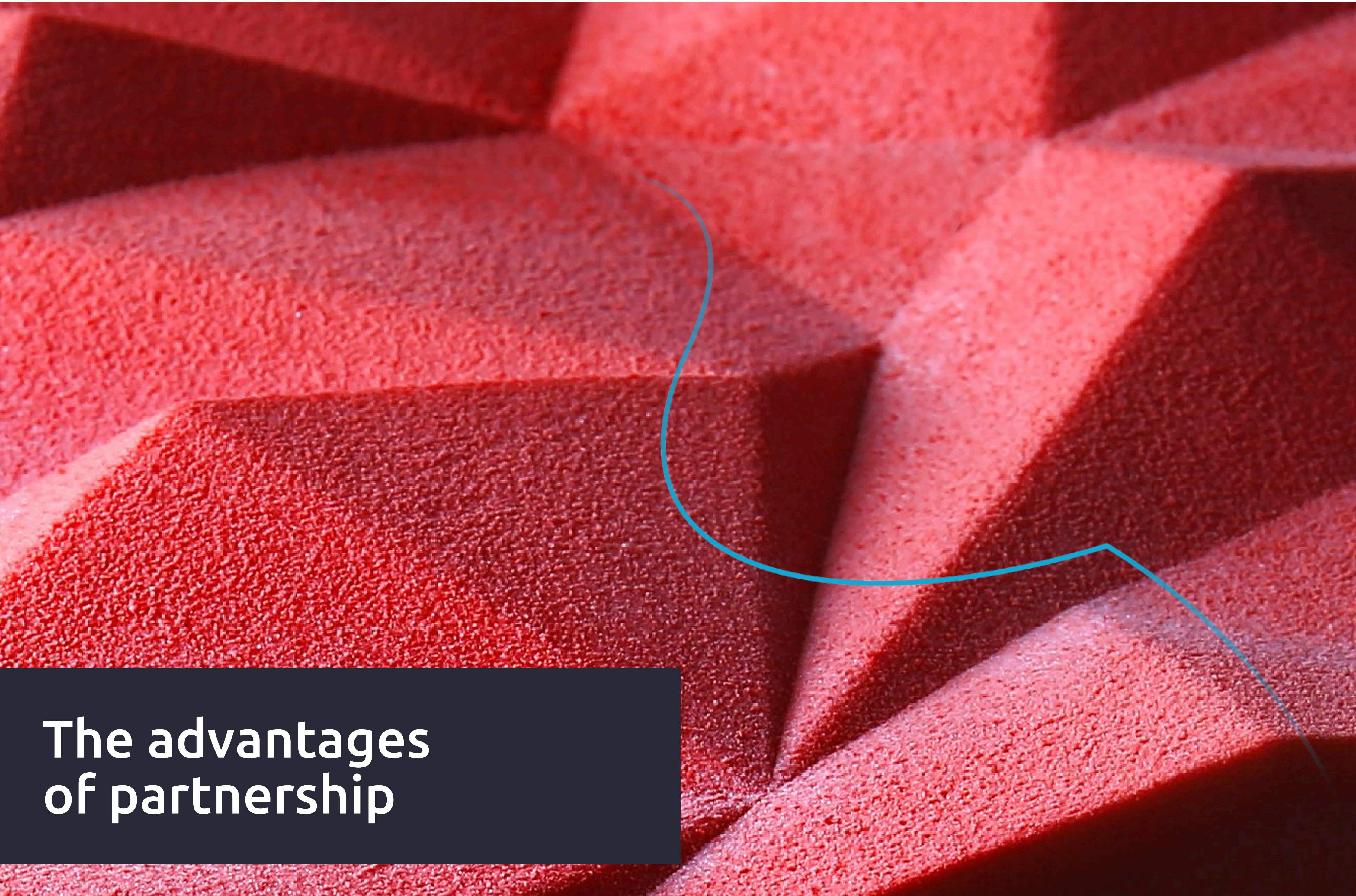
Architectural decisions will need to consider data sharing and distribution, privacy and security, the compute and storage facilities required, communication and connectivity, and load distribution.

The architecture also needs to address regulatory, legal, and safety frameworks.

Hybrid AI provides new tools to navigate this landscape and new options for creating engineered products and processes. The combination of generative AI and structured knowledge means that we can distribute information and activity across complex environments in ways that are unachievable by legacy approaches or monolithic generative solutions.

A set of principles like these will bring consistency, not only to the terminology used by engineering organizations, but to the approach they take. It will mean that, while Augmented Engineering solutions can be thought of custom 'vehicles', bespoke to the needs of the organization that deploys them, the wheels on which these 'vehicles' run won’t need constantly to be reinvented.



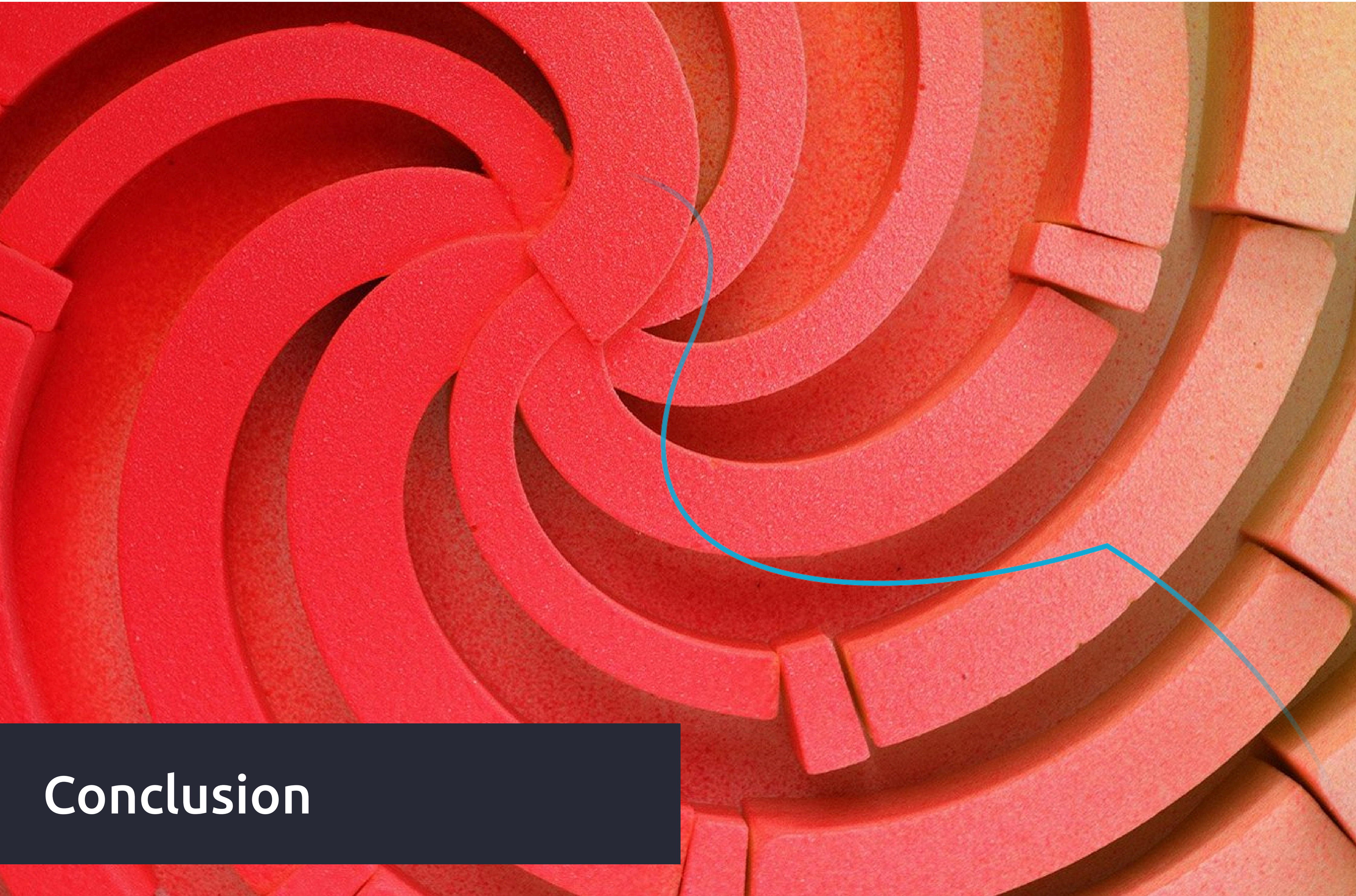


The advantages of partnership

Even if they operate on a global scale, individual enterprises tend only to have their own frame of reference. By contrast, a capable external services provider can bring knowledge and insight drawn from a wide range of experience, and can help to address challenges, even if those issues morph and grow.

Such a provider can contribute generative AI expertise in a number of ways, including:

- **Standardization:** Making the best use of emerging industry architectures and cross-sector experience for the best results
- **Enterprise:** Integrating into existing mature end-to-end industrialized processes, incorporating new AI-driven ethics, privacy, capability, trust, and enterprise governance factors
- **Commercialism:** Ensuring that Gen AI is implemented and managed in a hard-headed business manner that is mindful of cost and carbon boundaries, and that provides safeguards for privacy, intellectual property, and sovereignty
- **Ecosystem:** Established partnerships with the key hyperscalers and niche high-value players, sufficient to enable at-scale implementations
- **Compliance:** Explicitly managing the specific requirements of the engineering domain, including formally managed quality, compliance, sustainability, security, and safety
- **Capacity:** Demonstrable international Engineering Research & Development expertise to support organizations on their AI-enabled journey, as well as partnerships with major third-party developers, ensuring solutions can be deployed anywhere and at any scale
- **Engineering domain knowledge:** Clear strengths in software engineering, product and systems engineering, manufacturing and operations, and digital engineering
- **Innovation:** Close and productive relationships with leading commercial and academic AI innovators

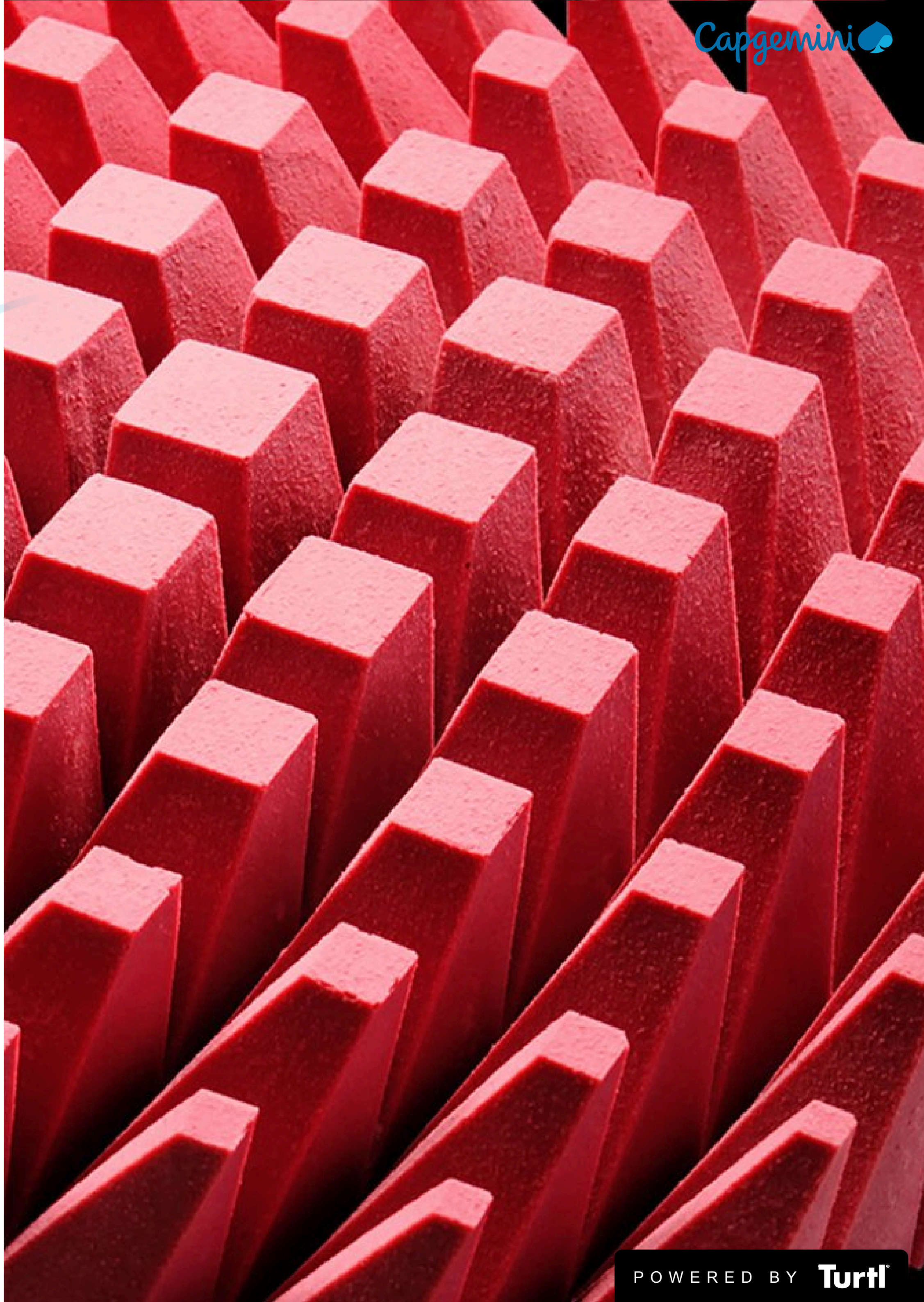


Conclusion

Artificial intelligence will undoubtedly change the nature of engineering. As we have seen, in this demanding discipline Gen AI needs to be incorporated into a more comprehensive and applicable Hybrid AI model.

The question for engineering leaders is how to plot the route to value creation for their organizations. There is no manual with simple steps that will lead you inexorably to success, but keeping front of mind the five principles set out in this document will make the journey smoother.

Andy Vickers
CTO Generative AI - Capgemini Engineering



www.capgemini.com