# Do not believe your eyes.

**The use of synthetic data for driver monitoring product development**

## A Seeing Machines White Paper

Author:
Timothy Edwards, Co-Founder & VP Strategy

Key contributors:
Lachlan Whichello, Senior Optical Engineer – Innovation
Alex Yee, Senior Staff Engineer - Data Strategy (Validation)

20 June 2023

**Contents**

# Introduction

For over 23 years, Seeing Machines has been engaged in the application and advancement of computer-vision technologies that perform real-time monitoring of vehicle drivers to prevent accidents. During this time, the field of applied computer-vision technology has evolved dramatically and, often, in sudden steps.

As the essential core technologies to implement driver monitoring have evolved and shifted, so have the engineering tools and methods for design and development, each step unleashing a host of new product possibilities.

The most recent of these shifts have been in machine-learning; in the model architectures, training methods and processor designs that together enable complex machine-learned models to be embedded in low-cost products. These have enabled a new wave of cabin-monitoring features previously thought unattainable at the commercial price-point for in-cabin systems.

On our journey to help society achieve “vision-zero”[1] we have also learned a great deal about the human beings that our technology is built to support. As our technology has become more able to detect and track humans, so has our understanding of behaviours that indicate potential safety risks. Recognizing the need for hard science in this area, we launched a human factors department in 2014 with the goal of guiding our engineers, customers, and regulators on the human aspects of driver monitoring technology.

Human factors science combined with the ongoing quest to improve our core technology places the discipline of data collection as a central core competency for our business. We have invested in test vehicles, laboratories and in partnerships with leading Universities around the world to ensure we have the best possible data to study occupant behaviour as it relates to real-world risk.

Additionally, our participation in the automotive market has created a deep awareness of the cost and timing pressures faced by the automotive industry and particularly how those pressures drive risk-based decision making throughout the supply chain. As a Tier2 supplier, Seeing Machines faces the same challenges, and we must continually weigh cost and schedule pressures against the performance we can achieve with each driver monitoring product we develop. Inadequate performance can mean either failing to protect the driver or annoying them with false alerts.

[1] https://en.wikipedia.org/wiki/Vision_Zero

At the intersection of the science of understanding human behaviour, a rapidly evolving vehicle safety market with cost and timing pressures, and the absolute need to increase real-world driver safety (ultimately reducing road accidents, injuries, and fatalities), we (like the industry at large) are continually seeking to further evolve our approach to Driver Monitoring design and development. One of the more controversial, elements in this evolution is the use of synthetic (or computer generated) data.

Synthetic data offers the potential to remove much of the slow and expensive real world data collection activities which dominate in-cabin product development efforts. Synthetic data can also theoretically provide access to data that may be nearly impossible to attain safely, or at all, by real-world recording means.

However, synthetic data also comes laden with risk. While the technology has been maturing in adjacent markets over the last decade driven largely by the global investment into automated driving, virtual reality and AI, its application to the human-centric world of driver monitoring remains in its infancy.

It is important that the both the benefits and risks associated with synthetic data are carefully considered. The adage "garbage in - garbage out" that holds true for other data sources absolutely applies here. Faulty or insufficient inputs will generate meaningless or erroneous outputs. Our internal studies have shown us that training our models with synthetic data can often result in models that fail to generalise to the real-world.

In this article we explore the application of synthetic data to our field of expertise; in-cabin monitoring systems. We discuss Seeing Machines' approach to the gradual and careful adoption of synthetic data in its internal engineering processes and why we have partnered with the specialist startup Devant.

## Introduction to Synthetic Data

Synthetic data refers to data which is created purely by software that attempts to simulate real-world inputs to any product. It is a powerful tool that is primarily used in the training and testing of machine-learned neural networks, otherwise referred to as “models”.

One of the earliest examples of the use of synthetic data in the automotive industry was in the development of autonomous vehicles. An early challenge of autonomous vehicle development was the extreme difficulty in collecting enough volume and diversity of real-world data. This is because autonomous vehicles must be trained to work in all possible driving conditions, including different weather, road types, and traffic patterns. To overcome this challenge, researchers developed synthetic data generators that could create essentially infinite driving scenarios. These generators required specialized simulations of road environments to be developed, to create data that attempted to be as indistinguishable from real-world data as possible. Despite limitations to the achievable realism, the use of synthetic data has allowed autonomous vehicle developers to train and test their models on far larger and richer datasets, leading to improvements that would not be otherwise possible using conventional means.

## Synthetic Data Market

Today synthetic data is used by a very wide range of industries, including finance, healthcare, security, manufacturing and of course automotive design. Accompanying this demand is a rapidly maturing sector of suppliers formed of companies entirely dedicated to the commercial supply of synthetic data.

There are many competing techniques for synthetic data generation. For a supplier to compete, they typically attempt to build highly flexible generation technology that allows their data to be adjusted to the needs of the widest possible range of different customers.

However, the problem of “jack of all trades, master of none” also applies. An alternative strategy is for suppliers to identify and develop expertise of niche high-value application domains, targeting customers who require exceptional quality and fidelity.

## AI Generated Data

While synthetic data is commonly used to train machine-learned models, machine-learning is also used in the generation of synthetic data itself.

For example, an AI product such as a chatbot might benefit from an infinite number of synthetically generated human conversations, which in turn could be generated by another AI that takes real conversations and blends them into new combinations of words or augments them with additional words.

Similarly, synthetic data for computer vision may incorporate AI models in many aspects of the “rendering pipeline”, including the scene generation, characters, animations and behaviours; as well as in deep inside the rendering technology itself, to improve ray-tracing quality, upscaling resolution, texture generation, or automatically creating “in-between” animation frames.

## The Concept of Digital Engineering

"Digital Engineering" is terminology for a method of engineering design that maximises the use of high-performance computing, data, simulation, models and automation, to unlock the power of *virtual and rapid iterative development*.

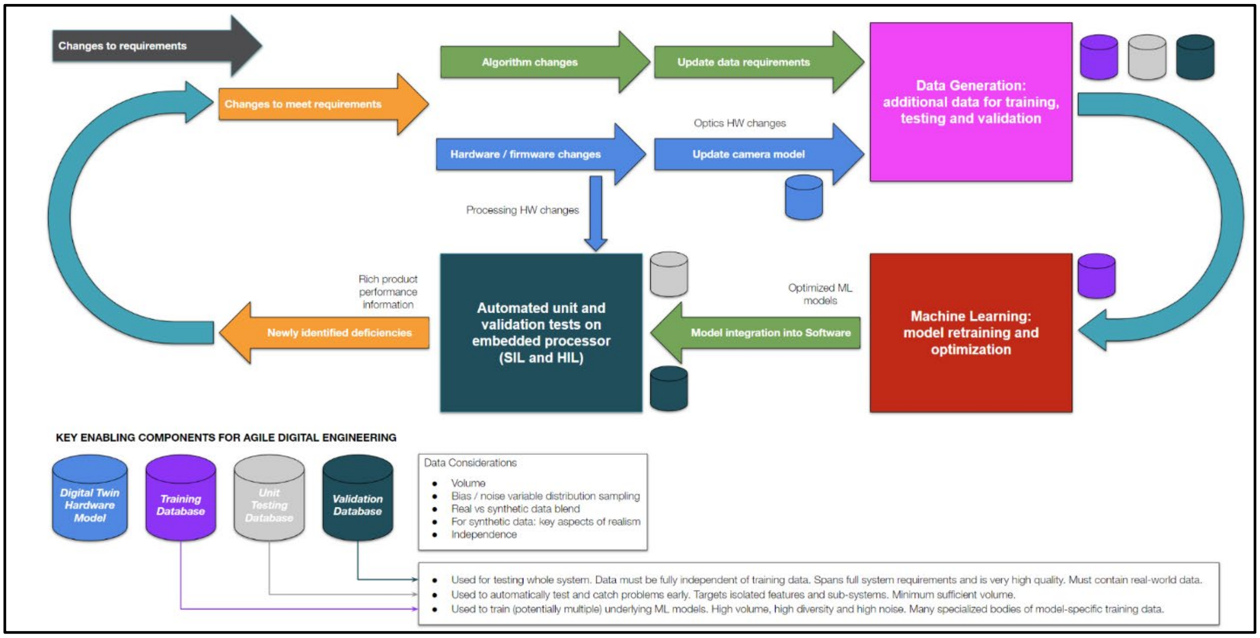
The general benefits include:

- **Increased efficiency**: engineers can utilise computers more efficiently by automating manual tasks.
- **Reduced costs**: costs are reduced by eliminating the need for physical prototypes on the innermost design loops.
- **Agility:** engineers have greater opportunity to alter or adjust the product design to meet its requirements by being able to quickly reveal the impact of changes.
- **Improved co-design:** engineers can explore changes to both hardware and software during the development, allowing hardware and software components to better fit together.
- **Improved product performance**: accurate digital models and simulations allow engineers to rapidly explore more design options, improving the probability of greater performance outcomes.
- **Increased safety**: simulations allow engineers to obtain a better understanding of potential hazards by allowing them to simulate and test designs against virtual safety hazards.

Synthetic data is a key enabling component for achieving Digital Engineering, particularly for any product that incorporates complex software algorithms (either developed conventionally or machine-learned). It may be used both for algorithm training but also to test virtual designs.

Of course, Digital Engineering has its limitations. Whilst fully virtual design and development is theoretically possible, for most modern products, the models and simulations always have practical limitations and there is always be the need for real-world testing and feedback in the design process. Digital Engineering simply attempts to reduce the magnitude of slow and expensive real-world testing, reducing the number of physical prototypes and manual test procedures that need to be developed to achieve the design goal.

## The Digital Engineering Loop




***Above**: a simplified view of an agile Digital Engineering product development loop. The goal is to remove as many manual steps as possible and minimise any developer friction around this loop.*

For cabin-monitoring systems, digital engineering requires:

1. Identifying the next change. This is most often a change to resolve an identified deficiency identified in a previous iteration but could also be a change to product requirements.
   - *For example, the software may be found to fail a performance KPI, such as eye tracking accuracy when the driver is wearing sunglasses.*
2. Adjusting any hardware or firmware to better mitigate the deficiency.
   - *To improve the product, the camera optical design, sensor exposure control loop or infrared lighting pulse characteristics, may be altered to fine-tune the impact of sunlight on feature performance. The camera location in the cabin may need to be slightly changed. This requires synthetic data systems that can accurately reproduce images of the cabin from any potential camera design under any scenario of lighting and occupant state, allowing engineers to make changes that are both linked to requirements and test-outcomes, while having a high degree of certainty as to the real-world outcome of the change.*
3. Creating data to better represent and reveal identified deficiencies.
   - *Testing may reveal that a feature performs poorly due to inadequate diversity, often in a corner-case (where several extremes from multiple "noise variables"*

*simultaneously occur). For example, a driver that is yawning while making a "lizard glance" down to their phone in very bright lighting conditions. Here, synthetic data may be used to quickly add thousands of targeted examples to the algorithm training set as well as new independent synthetic data examples to unit testing and validation databases.*

- *Testing may also reveal that the software does not run quickly enough on the target processing hardware, requiring harder optimization of the models, which may in-turn degrade their performance. In this case, the training volume may need to be increased to better fortify against any model degradation during optimization. Or where the processing hardware is also a design component, improving the processor design itself to more efficiently execute the models.*

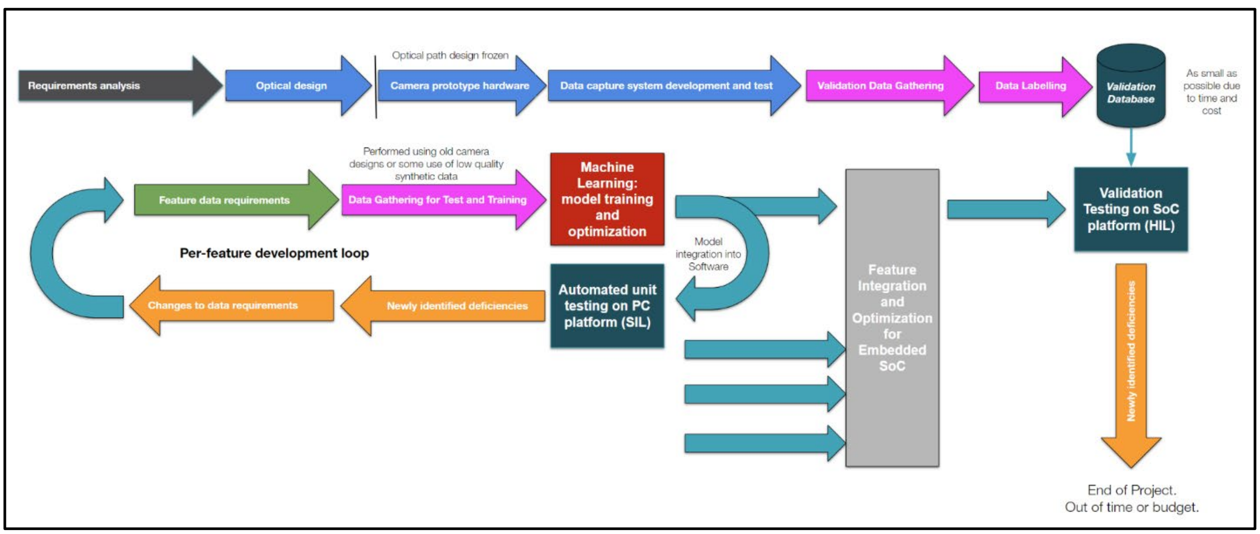
4. Retraining and optimising the affected models, reintegrating them into the software and running automated tests to confirm software correctness, via a continuous integration (CI) system.
    - *This is the complex world of "MLOps", which spans both traditional software engineering CI systems (DevOps) but adding the iterative loops required for embedded machine learning. These are automated systems that are supervised by machine-learning engineers.*
    - *DevOps (and MLOps sub-systems) can be especially complex as they need to automate multiple product development projects simultaneously, with version control of not only source code, but also databases and trained models.*
    - *Databases and models also become linked to bespoke product requirements including vehicle cabins and target SoCs. Each vehicle platform usually has a different SoC and contractually negotiated processing budget. This requires the creation of multiple different specialised databases and model variations, which must be continuously tested on each SoC target processor.*
5. Validating that the changes have reduced or eliminated the original deficiency by testing the full system against a database of carefully curated test data. This data must be *independent* of the training data to measure how well the models are likely to generalise to real world conditions.
    - *Here the validation data must be highly trustworthy, spanning all common combinations of product operating conditions. If the data is in some way deficient in realism, then product deficiencies are likely to be missed.*
    - *Achieving data independence with synthetic data systems is a key issue. One approach is to use different suppliers that generate data in different ways. However, that approach drives significant project complexities as managing each supplier, aligning the metadata and quality standards can take many months to*

*achieve. At sufficient realism, the better approach is to simply track and control the data elements, ensuring that the scene assets (including characters), cabins, cabin configurations, and all noise variables are always different between training and validation databases.*

- *Finally, because synthetic data will always have limits to realism, the backbone of validation data must always be some quantity of real-world examples. The ratio of real-world to synthetic data must be considered on a case-by-case basis and justification of this ratio requires a "data equivalence study" to be performed for each product feature.*
- *These studies must attempt to quantify how realism limitations may impact feature performance, particularly with regards to error distributions in output signals. Where it can be shown that synthetic data creates output signal noise levels which are similar (or larger) in magnitude to the real-world data, then an argument can be made for blending some synthetic data into the validation database (for the specific feature). The most common problem with synthetic data is the errors will be too small as the data are usually too clean, or in other words, they lack sufficient real-world noise characteristics.*

6. Studying the testing and validation outcomes and identifying any new deficiencies caused by the change in the training data or the hardware design, and if so, returning to step 1.

## Contrast to the Typical DMS Development Process




***Above:*** *A simplified view of the typical development process that occurs today. The validation data is very difficult to gather. The contribution of the validation database to the product is really only* *to identify deficiencies and there is rarely the opportunity to correct those deficiencies as these are only discovered very late in the project.*

Where synthetic data cannot accurately reproduce the system optics, the hardware adjustment step (2) must be skipped. In this case, optics hardware is designed at the beginning of the project and then the design “frozen” for the rest of the development effort. Here, late changes made to the camera design can invalidate the test or validation database. In our experience, it is extremely difficult to “freeze” a camera design early and minor changes to the optical hardware or firmware can and do occur continually during the project, leaving the Tier1 or OEM in the difficult position to really know to what degree they can really trust the validation database.

Another risk that may occur is whenever the synthetic data lacks sufficient realism with the effect on model training being an insidious reduction in performance. When this happens engineers must make educated guesses as to what aspect of realism the data is lacking. For example, perhaps a model for detecting driver distraction has automatically learned to weight decisions based on a combination of eye feature behaviours, face and hand movements, yet some aspects of these synthetically generated behaviours are subtly different to those in real humans. The model may appear to perform well when tested against synthetic data but underperform in the real-world.

Therefore, particularly for cabin-monitoring functions that have safety implications, synthetic training data must be used with extreme caution *or not at all*, and the latter has been the Seeing Machines policy for many years.

Finally, to pass validation testing (step 5), normally a video database must be collected by instrumenting vehicles with the “frozen” camera design connected to complex, expensive (and bespoke) video and metadata recording equipment. Actors (and directors) are used to induce behaviours that span all the possible conditions faced by the product, which may involve driving the vehicle on public roads, or on a test track, to help ensure data are truly realistic. Once the video recordings have been captured then manual or semi-automated video annotation must be performed. This is where the video is examined frame-by-frame to measure information needed for feature validation. This manual annotation process can be performed to some degree by automated algorithms, however there must always be a laborious expert review process required to check for errors, usually with multiple independent reviewers being required to ensure data is correct. In short, validation databases are very slow and very expensive to develop, and our experience is that the databases are often inadequate in diversity or size to reveal product deficiencies until the *very final stages of the development project*, leaving zero room for corrections should the validation reveal actual serious deficiencies.

*Obtaining a suitably diverse and correct validation database is one of the most severe risks in any cabin-monitoring product development project and to this day, is also one of the costliest aspects of integrating driver monitoring technology into any new vehicle*.

## The DMS Scale Up Problem

Seeing Machines first introduced vision-based DMS in 2017 to support hands-free partially automated highway driving systems such as GM's SuperCruise, and many other similar systems developed since.

Due in part to the success of these well-engineered systems (which have clearly delighted vehicle owners), and in part due to safety concerns over partially automated vehicles that do not use direct driver monitoring, we witness increasing demand and more ubiquitous deployment of vehicles with hands-free driving capability using vision-based driver monitoring systems.

However the safety benefits of vision based DMS are expected to be far more significant when applied to manual driving. This opportunity was recognized first by Euro NCAP. This year is a landmark year as it is when Euro NCAP commences a new procedure to test the performance of new vehicles sold in Europe on their ability to detect and mitigate driver distraction regardless of any driving automation mode. We anticipate that the industry-wide introduction of DMS will drive the societal change we have long sought after; where holding and using a phone while driving will be equivalent to driving without a seatbelt.

The Euro NCAP testing affects the star-rating of any new vehicle sold in Europe which has a powerful influence on customer purchasing decisions. The Euro NCAP approach is also backed by EU regulations that require drowsiness and attention monitoring to be in all vehicles sold in Europe from 1 July 2024, with similar regulations now in China and under discussion in the United States.

Since the ubiquitous deployment of vision-based DMS as a *standard safety feature* has only just begun, the "scale up" of vehicle-specific DMS products (at least in automotive terms) needs to be performed extremely quickly.

Additionally, beyond protecting road users against distracted and drowsy driving, DMS systems regulators are already considering several important improvements. Firstly, the states that define the distraction or incapacitation level of the driver are expected to become more refined. Examples include the definition and detection of an unresponsive driver. Clearly, with more data and a better understanding of the prevalence of additional driver states and the safety risks of those states, more enriched protocols are expected to be evolved.

In the years to come, we predict DMS systems will need to detect almost *all forms* of driver impairment including intoxication by alcohol or drugs or suffering an illness. The reality is, that once the most common conditions are managed, there will remain a very long tail of conditions that can still cause severe driver impairment and accidents. Examples include low blood sugar, low blood pressure, viral illness, heart conditions, severe indigestion, allergies, as well as forms of emotional dysregulation such as rage, panic or severe depression; anything that can impede a driver's ability to correctly perceive the road environment and operate a vehicle in a safe manner.

*In summary, the industry faces a nexus of issues:*

1. *DMS systems are trained to detect subtle human behaviours and operate optically in complex lighting environments. However, any change in key cabin geometries, camera location, or optical design may render existing DMS video test databases invalid, either fully or partially.*
2. *To meet regulatory timing, DMS must be simultaneously engineered into countless new vehicle models, but each major vehicle platform has a different cabin layout and camera location, with slightly different performance KPIs. OEM DMS supply contracts may span several Tier1 suppliers (or system integrators) having unique camera hardware designs, consisting of sensor, lens, lighting and exposure control firmware, as well as different processing hardware and compute capacities.*
3. *DMS driver safety features are experiencing continual ongoing development to advance the regulatory safety roadmap.*

*These issues combine to drive the risk that many DMS systems will become rushed to market; shoehorned into new vehicle designs with insufficient validation of their performance, leading to either (i) under-performing DMS that fails to adequately protect drivers in ways that a traditional testing approach is unable to reveal, due to the complexity and burden of creating suitable test coverage, or (ii) DMS systems that are poorly tuned and serve only to annoy drivers, leading to OEM brand damage or a general backlash to the safety technology as a whole, or (iii) both.*

*Faced with this scenario, it should not be surprising that the industry is interested in synthetic data as a means to reduce the cost and time required for DMS system validation.*

## The Validation Trap

As described above, system validation represents a significant engineering risk for every new vehicle platform.

First let us consider why validation is even needed. This question is not as ridiculous as it sounds. Most experts consider validation essential, but there is always the technical question; *what constitutes minimum sufficient validation evidence*? In hard commercial terms, “sufficient” is defined in terms of risk. Stakeholders across the industry, and at all levels of the automotive supply chain, must carefully consider the risk of being held partially or fully liable for a product failing to meet its performance requirements and that deficiency being directly linked to road-user injuries or fatalities.

Here it is important to understand that DMS systems are simply unable to perfectly detect and mitigate driver distraction or impairment. Their safety role is rather to materially *decrease the odds* of accidents. The exact same argument applies to airbags and seatbelts.

However even in this stochastic context, the performance of DMS products must be carefully crafted in engineering requirements that are negotiated between the car makers and their suppliers, usually as a set of measurable performance KPIs on each output signal.

Failure to meet any of these requirements has the potential to expose the supply chain to liabilities.

Hypothetically, in the event of an accident (or a pattern of accidents) where a driver monitoring system appears to failing to detect the condition it was intended to detect, directly leading to an accident event, *if it is also revealed that the DMS supplier validated the system using a video database based on a different camera, camera location or cabin layout that the actual production system, then a legal argument for liability may become stronger.*

*Consider further the sensitivity of a case where it is revealed that the validation database used to certify that system did not even consist of real-world recorded videos (let alone using representative system components) but was entirely simulated and computer generated.*

On the flipside, compared to conventional validation, it is no doubt true that synthetic data can be used to test a much wider range of conditions, arguably making any DMS solution more broadly validated and therefore theoretically more trustworthy.

Synthetic data also offers to save a DMS supplier a large amount of money, and reduce the engineering time by many months, creating powerful incentives to maximally adopt this form of data.

**The Validation Trap** occurs when the product supplier (Tier2, Tier1 or OEM) convinces themselves that synthetic data appears good enough to achieve the desired purpose and *does not do the hard work or accept the expense to prove that is really the case*. This may happen when someone mentally concludes that "*the data look real enough*", without adequate recognition of the fact that our human sense of "reality" is an illusion created by our visual cortex, that blends what our eyeballs see with what we think and feel.

Humans have evolved brains that are truly excellent at filling in the blanks, including in our own vision. We operate on patterns, not details. With computer graphics, our minds can be very easily fooled into believing that what we are seeing is an accurate reproduction of what is real. This is why special effects have been a successful industry for decades and why when you look back at an old sci-fi movie or computer game, the graphics usually don't seem nearly as good as when you first encountered them.

In short, in assessing the quality or fidelity of computer-generated synthetic data, it is extremely important to *not believe your own eyes*.

## Exploring Realism for Cabin Monitoring

In preparation for the Scale Up problem, Seeing Machines has been exploring synthetic data technology for several years.

In 2021, a Microsoft paper titled "Fake it Till You Make It" caused quite a stir as it showed the vast potential of synthetic data to create machine learned models that outperformed those trained by conventional means. The trick was the synthetic generation of an unlimited number of 3d face geometries, including expressions, which allowed the researchers to use rendering technology to create suitable artificial training data that broadly covered all the key variables of head-pose, eye-gaze, facial expression, lighting and items worn on the face, such as hats and sunglasses.

Compared to conventional means of training where 2d images or videos of people, or 3d face-scans, the researchers showed how synthetic data can be used to overcome the issue of *data scarcity*, in particular when synthetic data is used to *augment* real-world datasets.

While these findings were valid for the applications studied by the researchers, our own research found such techniques unfortunately could not be easily applied to driver monitoring system feature development.

**To harness synthetic data for driver monitoring systems, we are forced to examine its limitations**. We must ask ourselves, what aspect of the data matters and what does not? Or, to put it another way, what are the *aspects of realism* that are more important than others and can we *list and rank these aspects*?

The first step was to attempt to break down “realism” for driver monitoring, into a logical taxonomy to enable us to study each effect and to help us define a realism roadmap.

## A Realism Taxonomy

| Realism Class | Sub-Class | Component | Key elements |
|---|---|---|---|
| RENDERING FUNDAMENTALS | Scene Generation | Fitting of humans and objects into coherent cabin scenes, and animation of those scenes | Fitting of clothing and face props, moulding of humans into seats, [redacted]. Scene generation can be configured by [redacted] noise variables, their probability distributions and sampling of those distributions. |
| | Optical Correctness | Wavelength | Appearance of human body and clothing under near infrared vs visible light. |
| | | Lens | Lens FOV, distortion, [redacted] |
| | | Sensor and filters | Pixel response model vs wavelength [redacted]. Pixel noise model, [redacted] |
| | | Illumination | Light source 3d locations [redacted] |
| STATIC APPEARANCE | Human | 3D eyeball model | 3d model accuracy, [redacted] pupil [redacted] geometry [redacted] anthropomorphic variations, [redacted] |
| | | Eyelids | Shape variations, eyelash variations, [redacted] |
| | | Skin | Fitzpatrick classes, [redacted] |
| | | Faces | Age, sex, ethnicity, [redacted] |
| | | Glasses & Sunglasses | Frame shape and thickness, [redacted] |
| | | Other Face Props | Masks types [redacted], eye-patches, [redacted], jewellery |
| | | Head Props | Hats, caps, beanie, balaclava, scarves, and cultural garb (turban, hijabs) |
| | | Hands and hand-props | Skeletal model, anthropomorphic variations, [redacted] |
| | | Body shapes and sizes | Infants, toddlers, children, adults, [redacted] |
| | | Hair styles | Hair style variations |
| | | Facial hair | Shaved, light, bushy, [redacted] |

| | | | |
|---|---|---|---|
| | | Mouth, [redacted] | Mouth opening model, [redacted] |
| | | Clothing & Body Props | Cloth model, [redacted], fitting quality |
| | **Animal** | Body types | Cat, dog, [redacted] |
| | | Props | Collars, clothing, [redacted] |
| | **Cabin** | Trim Materials | Velour, neoprene, plastics, [redacted] |
| | | Geometry | Vehicle model geometric accuracy [redacted] |
| | | [redacted] | [redacted] |
| | | Main seat belts | Seat belt fit quality, [redacted] |
| | | Glass | Window positions, sunroof positions, [redacted] |
| | | Steering Wheel | [redacted] wheel materials |
| | | Infant carriers / Child booster / safety seats | Diverse range of infant carriers, [redacted] |
| | | [redacted] | [redacted] |
| | | Cabin Props / Accessories | Seat cover, Steering cover, [redacted] |
| | **Objects** | Hand held objects | Smartphones, [redacted]<br>Fit quality & natural hand poses |
| | | Placed objects | Backpacks, laptops, handbags, [redacted] |
| | **Lighting** | Environmental | Night, day, overcast, snow, rain, fog.<br>Large database of scene image variation.<br>[redacted] |
| | | In-Cabin | [redacted]<br>Interior lights on / off. |
| **MOTION ANIMATION** | **Human** | Eyelids | Blinking, [redacted] |
| | | Eyes | Gaze - Natural eye rotation [redacted] owl behaviour, [redacted] |
| | | Head pose | Movement with eye gaze (natural owl blend), [redacted], movement when speaking, [redacted] |

| | | | |
|---|---|---|---|
| | | Head command gestures | Natural nods and shakes of varying subtleness.<br>As per HMI requirements |
| | | Body pose | Various postures during normal driving, [redacted] |
| | | Hand pose | Various postures holding an object, [redacted] |
| | | Hand command gestures | As per HMI requirements |
| | | Hands on wheel behaviours | 11,12,1 o'clock, 9-and-3 o'clock, [redacted] |
| | | Facial Expression | Neutral, [redacted] |
| | | Mouth activities | Neutral, yawning, laughing, speaking, [redacted] smoking. [redacted] |
| | | [redacted] | [redacted] |
| | | Safe driving behaviours | [redacted] |
| | | Distraction (NCAP) | Specific behaviours described in NCAP as a reference test dataset |
| | | [redacted] | [redacted] |
| | | Drowsiness | [redacted] |
| | | [redacted] | [redacted] |
| | **Animal** | Behaviours & Pose | Sleeping, awake, [redacted] |
| | **Cabin** | Vibration | Smooth road, bumpy road, [redacted] |
| | **Lighting** | Solar | High sun, low horizon sun from any bearing, [redacted] |
| | | Night | Street lights, headlights |

***Above**: One potential taxonomy of realism for cabin monitoring (some elements are redacted for commercial sensitivity).*

What is important here is mapping the relationship between each realism element to any specific cabin-monitoring feature. Studies that examine the performance sensitivity to realism are recommended prior to adoption of synthetic data in the development of that feature. The map of relationships between realism and features is then used to guide synthetic data R&D efforts.

# Synthetic Data Technologies

This section offers an overview of some of the technology available to synthetic data suppliers and describes some of the advantages and limitations of each approach for use in driver monitoring systems development.

## Character Generators

The ability to synthetically generate unique character bodies and faces is a very important issue for the development of cabin monitoring products because these are products that are focussed on supporting and protecting people in vehicles and they must therefore exhibit the least possible bias in their performance for any individual.

The first face generators were developed in 1997 through a method called Active Appearance Models (AAMs). AAMs are generative models that can be used to generate new faces that are of similar appearance to faces in a training set. Face AAMs are built from statistical shape models that distil the shape and appearance of human faces (discovered from a set of images of faces or in some cases, full 3d scans of faces), into their principal components (or dimensions) and characterising the distributions of the data in the dimensional space. Importantly, faces generated by an AAM will have shapes that fall within the distribution of the real faces that comprise the training set.

The same approach may be applied to body shape generators, where the input databases consist of 3d scans of bodies of various shapes and sizes.

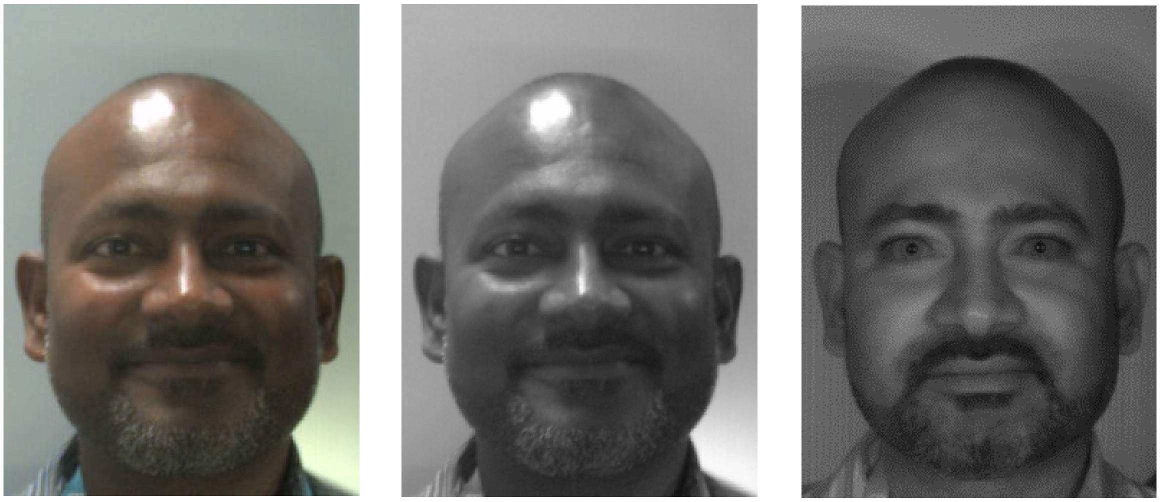
The quality of these generated models is only as good as the databases behind them and therefore the databases that underpin the models are considered commercially valuable and are therefore hidden from inspection. This means that for most character generator technologies, it is not possible to examine for bias in age, sex, height or ethnicity.

Additionally, appearance models account only for the geometric shape of the face and body. To improve realism, character generators also need to generate:

- Skin. Skin is a very complex biological structure that has varying transparency to different wavelengths of light. Near infrared light undergoes lower absorption and scattering than visible light, primarily due to the spectral properties of melanin. This gives skin a markedly different appearance when viewed with a near infrared camera compared to what one might expect based on a visible light camera. The deep penetration of near infrared, combined with subsurface scattering from the skin layers and venous structures, creates a “soft glow” effect that emanates from the skin.

- Eyes. Eyes are internal structures (and are part of the brain), which are not modelled by AAMs. All face generators we have evaluated, assume a fixed model of the human eyeball, when in reality the geometric dimensions of human eyes have their own set of natural distributions. Anatomical data obtained through MRI scans and other methods, to some degree, can be used to capture eye structure variations. For high-quality driver monitoring where the accuracy of the eye-gaze direction is a very sensitive system variable, it is vital that the natural variation in adult eye-structures is accurately represented in any synthetic data. While these variations may seem small, in some cases spanning only a few millimetres, they play a significant role in eye-tracking accuracy. Seeing Machines has developed its own multivariate statistical model of eyeball geometries which we apply to our synthetic data generation efforts.
- Hair. Head and facial hair can drastically alter the appearance of a face, and facial hair that falls over the side of the face or which interferes with the view of the eyes, can directly impact driver monitoring performance.
- Animations for facial expressions. To animate the face and body, the underlying facial muscles and bones must be modelled. For driver monitoring, the correct animation of the eyelids and eyes are of particular importance, and this remains an ongoing area of research as synthetic data systems only offer basic eyelid blinking behaviours. Blinking events have timing and velocity distributions and impaired drivers in particular, exhibit changes in their eyelid and gaze direction movements which can be used as a clue to their cognitive state. Similarly, subtle eye motions may be used to differentiate between short glances down to a phone vs microsleeping, which are visually quite similar (yet different) motions.

***Right****: Example of [Unreal Metahuman](), which is a service that generates 3d animatable models for use in computer games. The characters appear visually realistic to the human eye, but these models lack validity for use in driver monitoring product development (for example for DMS features that are underpinned by gaze tracking, facial expression or any features that involve complex and subtle behaviours including distraction, drowsiness, intoxication, emotional states or sickness).*



***Above left***: *Fitzpatrick skin type 5 under visible light.* ***Centre***: *The same image transformed to grayscale by taking the luminance channel.* ***Right***: *Exposure under near infrared light creates a very different image and this can have a significant impact on driver-monitoring performance.*

## Renderers

Renderers are software components that consume 3D models of scenes, and which generate images from them. A renderer computes lighting, shadows, and material reflections in the scene, and then uses this information to create an image.

### Physically Based Rendering

Physically based rendering is described by the landmark textbook [Physically Based Rendering, From Theory to Implementation] authored in 2014 by Matt Pharr, Greg Humphreys, and Pat Hanrahan. It describes the mathematical techniques required to model the physics of light as it interacts with physical objects.

***Above**: human skin model that simulates sub-surface scattering of visible light, refracted through 7-element fisheye lens assembly. This was rendered in PBRT which is an open source "reference" renderer developed alongside the textbook[2].*

In the past renderers differed greatly in their fundamental techniques. However today, all modern renderers use these "physically based" rendering techniques, implementing the methods described in this textbook.

### Ray Tracing

All modern renderers employ the technique of ray tracing to simulate the way light reflects and refracts in a geometric scene. As the name implies, it is a technique that traces the path of light rays as they travel through a scene, from light source(s) to the camera. The technique supports rendering highly realistic images with accurate lighting, shadows, and reflections.

[2] Credit to Laubwerk (Jan-Walter Schliep, Burak Kahraman, and Timm Dapper) for the landscape, and Infinite Realities, Inc. for the humanoid.

Ray tracing is a computationally expensive process, but it has become more affordable in recent years due to graphics hardware (GPUs) designed specifically to support the ray tracing algorithm at the silicon level.

### Material Properties

Ray tracing as a technique does not completely simulate photon interactions with objects, rather it simply explores the pathways that light can take through a complex geometric scene. Much of the realism in renderers boils down to how the ray tracing is integrated with mathematical models of light reflection, refraction and scattering. These mathematical models are implemented as different “materials” which can be applied to objects in the scene. For example, one material may simulate a metallic surface, reflecting nearly all the light without any scattering, another may simulate a rough fabric, absorbing some rays and scattering others in a distribution.

The sophistication of material models and the size of the libraries available for animators to use, is a point of difference to consider in renderer technologies.

All renderers operate with libraries of materials that model only visible light wavelengths.

*There is no known material library that correctly models common materials at infrared wavelengths.*

### Game Renderers


***Above**: Unreal Engine 5 represents the state-of-the-art for computer games, offering superb visual realism at high framerates on a desktop GPU.*

Game renderers are optimised to obtain the maximum framerate from GPU hardware. As such these renders offer very *low-cost* synthetic data. Quality is also improving all the time driven by the fierce competition in the entertainment industry.

The key issue with game renderers is they are engineered very deeply to prioritise real-time performance above all else. In practical terms, this places limitations on customization and in our experience, suppliers that use these engines will draw firm lines on what they can and cannot deliver in terms of achievable realism.

### Commercial Renderers

***Above**: VRAY by [Chaos](#) is an example of a renderer commonly used in architecture and product marketing. Competing products include Arnold, Mental Ray, Octane and Redshift, to name a few.*

Compared to game engines, commercial offline rendering technologies do not need to strictly target real-time performance. The users may be architects or film animators, who can accept longer render times to achieve the levels of realism they seek. When maximum realism is needed, the scene complexity, materials and ray-tracing drive memory demands beyond the limits of GPU hardware. Therefore, unlike game engines, these types of renderers incorporate sophisticated distributed computing back-ends, allowing the calculations to be spread across large clusters of CPUs.

## Animation and Motion Capture Realism

Cars and people are things that move. Accurately animating movement is a key topic of realism that must also be carefully considered for in-cabin product development.

For example, an eye-blink may seem like a simple enough animation. The eye simply closes and then reopens. However exactly how eyes blink, including the wide distribution of timings and velocity across the human populace, is not so easy to describe. Additionally, how eye-blink behaviour changes when someone is drowsy due to lack of sleep is its own research topic.

The same fidelity issue applies to most aspects of human movement. As humans we have evolved social survival mechanisms that require the precision assessment of the movements of others. Our brains can quickly determine if someone is unwell, injured, excited, bored, etc, all just from the various movements of their body parts. For this reason, film animators appreciate that one of the most sensitive and important aspects of realism to convince a human audience, is the movement of faces and bodies.

The highest quality animation of humans is obtained through the use of specialised motion capture systems, which incorporate hardware that is worn on an actor's body. While face and body motions can certainly be captured without such hardware, to this day there remains no substitute for the accuracy and signal fidelity these systems offer.

***Right****: motion capture systems use custom hardware to ensure high accuracy 3d tracking of key points on an actor's face and body. These points are then used to drive an underlying model of bones and musculature. The model state is then transferred onto an equivalent model that animates the virtual character (which may have very different face and body proportions).*



seeingmachines

### Motion Capture Libraries

Motion capture systems are used to by animators to capture components of animation sequences (for example, a warrior thrusting with a sword). Groups of sequences are then stored in motion capture asset libraries. Animation software enables animators to create composite animations which may be blended from multiple motion squences (for example, a warrior that thrusts, defends, and dive rolls, all as one smooth sequence, formed from three separate motion capture assets).

Animation software also offers algorithms that enable animation sequences to be adjusted to fit the scene. For example, a character's motion may be scaled up or down in magnitude, resampled in time (to slow the animation up or down), clipped, limited or bent in order fit other geometry in the scene.

Finally, generative AI can potentially be trained on a library of similar motion sequences, and used to generate infinite new novel sequences which are similar in nature to those in the library.

### Animation of Impaired Drivers

The safety features of in-cabin monitoring systems are concerned with detecting potentially hazardous behaviours of the occupants, with a key subset of hazardous behaviours being those driver behaviours that indicate the driver's abilities may be impaired in some way.

The most obvious example here is driver distraction.

But what defines distracted behaviour and so how does an officially distracted driver really behave?

Working with industry groups and Euro NCAP, Seeing Machines has helped define a suitable set of high-risk driver behaviours that driver monitoring systems ought to be able to detect. The behaviours are largely described by head and eye movements, as well as hand activity, such as holding a phone, or the steering wheel. While the Euro NCAP approach calls for some relatively extreme examples of these behaviours to be used for testing driver monitoring functions, Euro NCAP also requests that car OEMs provide a body of evidence (or "dossier") that demonstrates how effectively their vehicle detects and mitigates driver distraction across a broad range of conditions.

It is important to recognize here that despite the concise behavioural descriptions for distraction published by Euro NCAP, there remains room for an infinite number of potential variations in driver behaviour that can either be considered distracted, or not distracted.

While some driver monitoring solutions will detect distraction based on measurements of head and eye gaze angles, timing and thresholds directly derived from the Euro NCAP test protocols, Seeing Machines knows that to avoid unnecessary false positive warnings (risking driver annoyance), sophisticated solutions will need to employ algorithms that are tolerant to the many subtle variations in distracted driver behaviour.


***Above**: Example image from a synthetic data sequence showing a driver becoming distracted by their phone.*

## Camera Simulation

Cameras used for monitoring vehicle cabins are highly engineered devices, optimised to the specific functions they support.

In-cabin cameras must operate over extremes of temperature and lighting, delivering suitable quality images to support the vehicle's functions under all conditions that the vehicle is intended to operate.

Due to cost pressure, in-cabin cameras are also designed to be “just good enough” to meet requirements and typically deliver images with lower resolution and visual quality to cameras in mobile phones.

More recent camera designs simultaneously operate in both visible and infrared wavelengths. For infrared wavelengths the cameras employ active illumination to improve the exposure conditions in the presence of both bright sunlight as well as for very dark conditions. One problem is that image sensors have relatively low quantum efficiency at infrared (this is the efficiency of the conversion of photons to electrons), which drives the need for active infrared illumination to be applied. Similarly, infrared light source components are poor at converting electricity to infrared photons with about half of the input power being converted to heat. These issues combine to cause thermal management to be an important design consideration, leading to complex trade-offs at the system level.

One approach is to ensure the infrared lights are only active when the image sensor is collecting light (during the shutter period). To minimise the overall time where the LEDs are active the image sensors must employ a “global-shutter” approach whereby all pixels in the array are exposed at the same time. Additionally, exposure control algorithms dynamically adjust both sensor properties (shutter period, amplifier gain) as well as LED activation intensity and timing.

Of course, camera design solutions used in the industry are by no means fixed and suppliers of cameras and their components compete fiercely on price and performance. Consequently, new techniques and technologies are continually being introduced by sensor, lens and lighting suppliers. Incorporating candidate technologies into the camera system and evaluating the impact, is slow and laborious work.

Simulation of in-cabin camera systems is presently not available to product designers and represents a gap in the synthetic data market. While training data can be carefully designed to side-step these issues, *it is clear that an accurate simulation of the camera is a requirement for synthetic data to be eventually accepted for validation purposes.*

***Above**: Visible light render of cabin scene, without any camera simulation.*



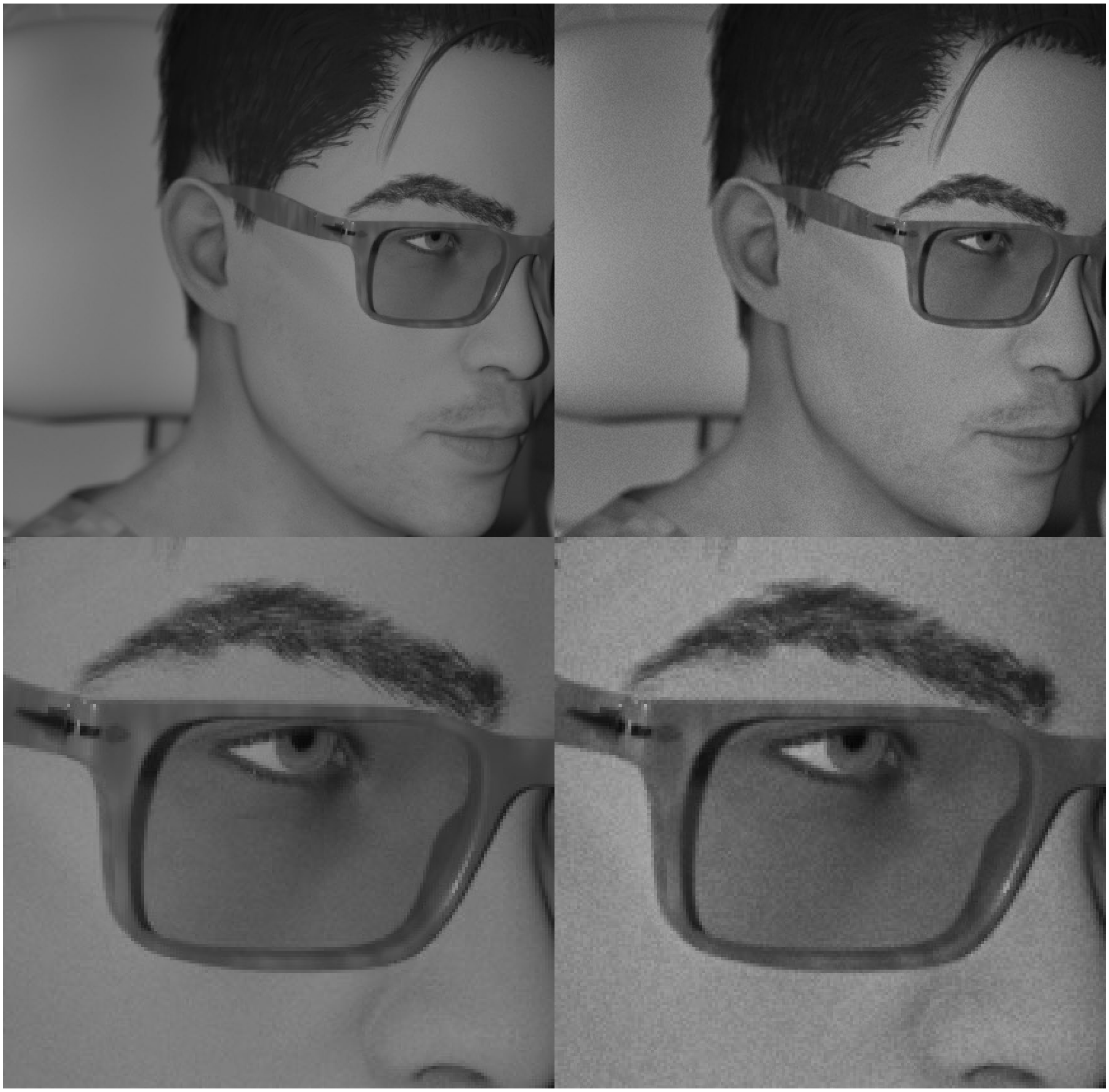
seeingmachines

***Above:*** *Infrared wavelength render of same scene.*

***Above***: *Active infrared illumination model applied.*

***Above:*** *Lens model applied. Here a fisheye lens with finite aperture is used. The image now has accurate distortion, depth-of-field blurring and attenuation effects.*

***Above:*** *Sensor model applied. In this final stage of the camera simulation, the image is sampled at the resolution and bit-depth of the image sensor, with sensor-specific and temperature dependent noise properties.*



***Above:*** *Sensor model before and after zoomed to reveal the noise. In this final stage of the camera simulation, the image is sampled at the resolution and bit-depth of the image sensor, with sensor-specific exposure and noise properties. The noise sources must be highly accurate in order to co-design the camera and algorithm solutions as a system.*

## Generative AI

AI "diffusion models" have shown powerful capabilities to generate unique and novel text and images and to some extent, video. Services such as [Stable Diffusion](Stable Diffusion) demonstrate the power of these large models to hallucinate imagery from the simplest of inputs, such as a verbal description of an image.

The state-of-the-art in diffusion models is moving very quickly and exactly how this technology could be wielded to augment the conventional “model and render” approach will be interesting to observe.

For cabin-monitoring, the direct benefits are expected to be the generation of novel cabin geometries, interior trims and textures as well as automatic creation of novel human appearances, clothing, behaviours and random objects in the cabin. For example, how can a cabin-monitoring system be built to detect a child-seat that has not even been designed yet? The answer may be to use a diffusion model to hallucinate a vast number of potential child seat designs and train the models against this kind of data.

However, there are many potential pitfalls and issues with using generative AI this way. Firstly, the synthetic “truth metadata” must also remain accurate to the generated image. Additionally, diffusion models are known to be hard to control and there is always the risk that somewhere during the generation of millions of frames of synthetic video, the model will generate something unacceptable. This may not be a major problem when generating training databases, but erroneous data is obviously unacceptable for validation.

With the use of these techniques, there are also commercial risks, because machine-learning engineers will create derivative IP through specialisation of an existing model or “backbone” already used in the open-source community. The background training data and rights can be murky in these circumstances.

This is a minefield, and Seeing Machines is exploring this space very carefully. However, we do expect diffusion models to find their way into rendering pipelines at certain key stages to create carefully *controlled novelty*. However, we do not see traditional rendering technology being replaced by AI, rather a hybrid approach is likely to be the next step.

***Above left**: synthetic image of vehicle cabin obtained from a data supplier (not Devant). **Right**: improved and novel realism of character appearance in the same cabin with identical body-pose. However, in this example the eye-model was corrupted, illustrating the need for highly controlled and careful application of diffusion models in the synthetic rendering pipeline.*

## Seeing Machines Partners with Devant

As described previously, the “scale-up” problem for the in-cabin industry is becoming increasingly complex, driven by the fact that:

- Each OEM may engage with a range of suppliers with varying hardware solutions, with different sensor, lens, lighting, processing, and electronics packages.
- Each OEM vehicle platform has a different cabin geometry and camera location(s).
- Each OEM seeks to differentiate, with unique feature ideas and requirements.
- Safety features and testing methods and are continually evolving.

The standard approach to managing this complexity is to develop a general-purpose software API that is maximally agnostic to system hardware differences and attempts to generalize the common requirements for all possible hardware combinations. While this approach is a good way to limit development costs, as the number of system design variables grow, the approach risks creating mediocre solutions that fail to delight customers.

As an alternative Seeing Machines is exploring an agile Digital Engineering approach, seeking a more efficient means to enact the algorithm-optics-processing co-design process to be able to offer specialized solutions that meet all customer requirements. The key to this approach is machine-learning process automation married with an efficient synthetic data pipeline that is able to simulate customer-specific camera designs and cabin layouts. Ultimately, we aim to deliver richer feature sets and enable OEM differentiation, without incurring additional development costs.

In our quest to develop this pathway it quickly became clear that we required a synthetic data technology that was 100% targeted at in-cabin product development. While there are many synthetic data suppliers who are seeking to service this niche market, no party has been able to offer Seeing Machines sufficient realism or configuration flexibility necessary to achieve our goals.

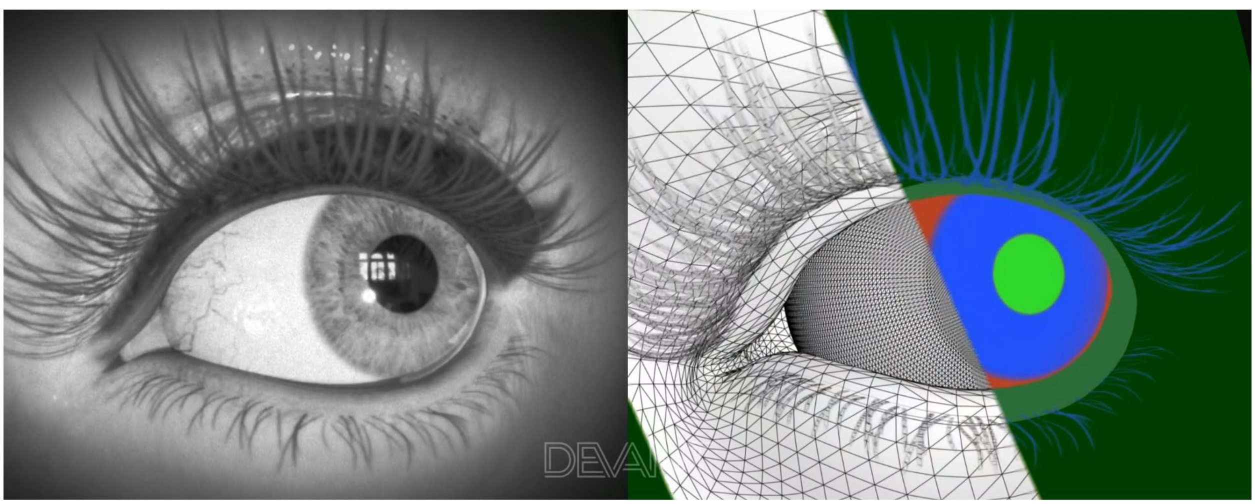
We, therefore, sought a partner willing to co-invest in our vision and innovate alongside us. This had to be a party that was willing to incorporate our insights in driver behaviour and technical knowledge of in-cabin products while also pushing the limits of synthetic data generation in terms of the quality, fidelity, and efficiency needed to unlock the Digital Engineering approach we seek for DMS and future in-cabin monitoring applications. That partner is Devant.

Devant is a Swedish startup, founded by film and animation industry veterans. Devant’s ambition was to work with a domain expert such as Seeing Machines to develop a best-in-class solution, starting from the ground-up to achieve the highest possible quality and efficiency.

Using the latest commercial rendering and pipeline technologies, motion capture systems, and with a team of skilled software engineers, animators and machine-learning engineers, Devant have built a fully automated pipeline with a configuration system that is wholly targeted to the realism aspects and noise variables of concern for cabin-monitoring products.

The collaboration between Seeing Machines and Devant is enabling Seeing Machines to use synthetic data to quickly explore new features as well as greatly improving the performance of our existing models. Together with key customers, we are also mapping the relationship between the realism taxonomy and key system features, to unlock the use of synthetic data for system validation and therefore to achieve our agile digital engineering goal[3].

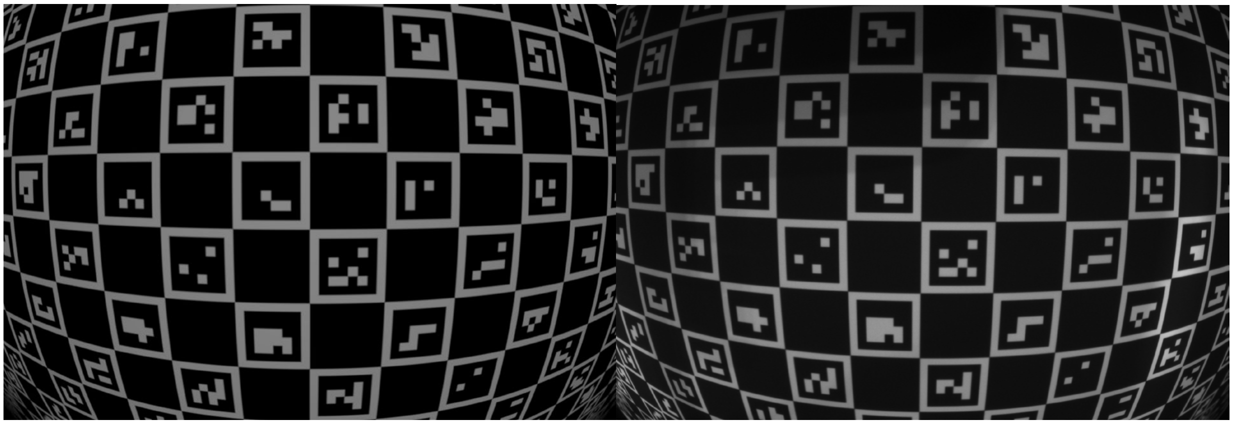
This work also forms part of a larger company goal to ensure car OEMs can measure and compare the quality of different cabin monitoring products, particularly with regards to safety performance. This is an issue that will become increasingly important over the next few years as in-cabin products become commoditized and price pressure intensifies.




***Above:*** *Eye model developed by Devant to meet Seeing Machines requirements. Anthropomorphic dimensions (and variations in those dimensions) are derived from Seeing Machines twenty years*

[3] Please note: Seeing Machines presently works with multiple existing data suppliers to obtain manually gathered (and AI labelled) data for use in algorithm training, automated test and verification systems. To this day, our company policy remains to use only real world data to develop our production technology. As this paper describes we are presently studying where synthetic data may be applicable to augment and improve training, testing and validation processes.

*of eye-tracking research. Additionally, the correct animation of the eyeball, eyelid and the pupil is a key aspect of realism for some cabin-monitoring features.*



***Above:*** *Accurate optical simulation is a key requirement and lens distortion plays a big role.* ***Left:*** *Synthetic reproduction of optical target,* ***Right****: real laboratory image of optical target.*


***Above:*** *screenshot from a face-tracking equivalency study.* ***Left****: infrared video of a real driver wearing a motion capture rig.* ***Right:*** *synthetic reproduction with the same head-pose, eyelid, gaze and facial expression data on a virtual driver. This study found that the synthetic data could certainly reproduce eye-tracking performance in benign conditions but still lacked sufficient noise characteristics to be accepted as validation data.*

## Devant NCAP Validation Service

The "scale-up" problem exists not just for Seeing Machines and its competitors but impacts the whole of the DMS (in-cabin monitoring) supply chain, including silicon suppliers, optics suppliers (camera and illumination), electronic control unit suppliers, system integrators, and OEMs.

While synthetic data offers a path to reduce the cost, complexity and time for validation, we caution that synthetic data on its own should not be regarded as the silver bullet that magically solves these challenges. Do not believe your eyes.

Rather, the careful use of *high-quality and carefully targeted* synthetic data offers a way to reduce the time, cost and *risk of validation failure* that could otherwise occur when:

- developing a new cabin-monitoring product from scratch;
- adjusting an existing product to fit into a new vehicle model;
- exploring changes to reduce the cost of the camera or processing hardware.

To facilitate this goal, Seeing Machines and Devant are working towards building the gold standard synthetic data platform to assist OEMs in their efforts to validate that their new vehicle designs will pass NCAP tests and meet the various regulatory requirements, as quickly and painlessly as possible.

Recently, Seeing Machines' knowledge of the fundamentals of driver distraction behaviours have been applied in Devant's motion capture studio, to create a high-fidelity motion-capture database. This database serves as a key input for generating endless examples of distracted driver behaviours that fit within the NCAP definitions. We anticipate that animations generated from this database will be a powerful tool for vehicle OEMs to measure DMS performance for both true and false positive conditions.

| Aspect | Typical Validation Database | Devant Service |
|---|---|---|
| Camera | Camera design and location must be frozen before data is collected. Custom data capture hardware requires fit-out to special test vehicles. | Multiple virtual camera designs and locations can be explored at the same time. Production camera designs can be characterised and accurately simulated. |
| Vehicle cabin | Recordings from an older vehicle model which may have differences to the production model. | Can be exact geometry of production vehicle(s). Multiple model variants and cabin trims can be validated at the same time. |
| Test subjects | Hard to recruit appropriate diversity in height, age, physical appearance. Trained instructors needed to direct driver behaviour during recording. | Balanced database diversity for subject size, shape and appearance, including skin colours, clothing and items worn on the face. Behaviours animated using pre-existing motion-capture library. No subject privacy issues. |

| | | |
|---|---|---|
| Noise variables | Can only explore limited combinations, cannot fully examine or discover corner cases. | Configuration system that allows for control over sampling of known noise variables across their probability distributions. |
| Test Compliance | Fixed to test requirements known at time of recording. | Flexible to changes in test requirements. |

***Above****: Image from a visible light animation sequence where the driver is reaching to the front passenger seat.*

***Above***: *Images from infrared animation sequence where driver is reaching into the rear footwell. This is a high-risk distraction behaviour.*

## Contact

Devant is presently seeking expressions of interest in a synthetic NCAP validation service. Contact Richard Bremer, CEO Devant, richard@devant.ai