WHITE PAPER

Machine learning for materials design and development

Cut development times and find new solutions for alloys, composites, and other advanced materials

intellegens.com | info@intellegens.com
Intellegens, The Studio, Chesterton Mill, Cambridge, CB4 3NP

intellegens

Applied machine learning

Executive Summary

Materials development is fundamental to the global economy and to net zero goals. This is true both of R&D by primary providers of **metals**, **plastics**, **composites**, and other **advanced materials** and of materials and process development in manufacturing industries such as **aerospace**. There is a continued need for fast, agile innovation to develop speciality capabilities that can drive value in commoditised markets and to address challenges including supply chain and price volatility, environmental and regulatory targets, and maximising the value of corporate IP. Achieving speed and resilience requires teams to understand complex, high-dimensional, non-linear relationships in materials systems in a context where property data is often incomplete. The result is high reliance on costly experimentation.

Machine learning can help by learning from experimental and process data to create models that capture these high-dimensional, non-linear relationships. Such models can then be used to explore design space for materials and processes, to identify potential solutions, and to focus experimental work. But machine learning methods can struggle where data is sparse, as it often is in materials R&D. These methods are also difficult for domain experts to apply and interpret. And their use is constrained by concerns about trustworthiness and transparency.

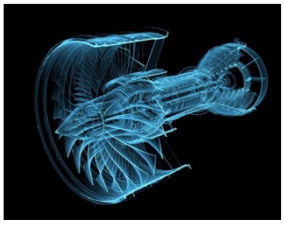
We introduce **Alchemite™**, machine learning software that overcomes these difficulties. We share application examples in which Alchemite™ has been used to provide valuable insight, optimise products and processes, and reduce experimental workloads. These include projects in steels, aero alloys, additive manufacturing, and composites. Polymers are covered in a separate Intellegens white paper.

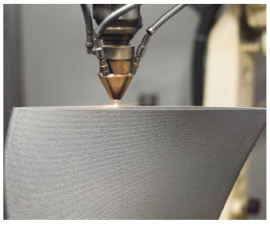
Materials and processes – the challenges

Strategic importance

> Materials and process development is fundamental to the global economy and net zero goals

The development and optimisation of materials and related processes is fundamental to the global economy. Economically, metals remain the dominant material type, with a global market estimated at $4 trillion [1]. But the plastics industry was worth over $600bn [2] in 2022 and composites over $110bn [3]. What these figures conceal is the additional scope and importance of material development work within organisations whose primary business is not selling materials – e.g., aerospace, automotive, or battery manufacturers. For example, novel alloys and composites are fundamental to the future of the aerospace industry, with composites now comprising 50% of materials in major aerospace programmes [3]. Successful materials and process R&D is thus a strategic imperative for global manufacturing, to ensure more performant products, to lower costs and development times, and, perhaps most significantly, in achieving net zero goals. This latter objective is reflected both in the need for new materials that can make products less environmentally harmful – for example, by enabling lightweighting of vehicles [4] – and in the need to reduce the impacts of material production – varied estimates find that steel production alone is responsible for 7-11% of global carbon emissions [5].






Commercial challenges

> Achieving competitive advantage requires specialist products and services

Materials R&D teams, whether working for primary materials producers or in industries that depend on materials, such as aerospace, need to help their businesses meet some key commercial challenges:

The innovation imperative – for many raw materials, particularly metals, markets have become commoditised, with prices driven down by global competition and economies of scale. See, for example, China's dominance of world steel markets [6]. To find competitive advantage in such markets and drive-up margins requires development of specialist products and related services. In the case of steels, this means green steels or steels that can fulfil specific customer requirements through unusual combinations of physical, mechanical, thermal, and corrosion properties.

Volatile costs and supply chains – making a material requires raw materials and energy as inputs. The current global political situation has resulted in significant fluctuations in the availability and cost of these resources [7]. Materials R&D needs to be increasingly nimble in responding to such changes, for example, by identifying alternative raw materials or proposing lower energy processing options.

> Materials R&D needs to be increasingly nimble in response to market and regulatory changes

Environmental and regulatory factors – almost every corporation now sets targets to lower the environmental impact of its products and processes. Government regulation is increasing, not only to achieve net zero, but also to reduce usage of materials and chemicals that have negative effects on human health or the natural environment, for example, through measures such as the Reach regulation [8]. For materials R&D teams, these factors add additional constraints to what are already complex multi-dimensional optimisation problems.

Intellectual property – materials science is an area in which R&D organisations remain highly-dependent on the intrinsic knowledge of experts. While this human factor will always be important, and in fact is something that needs to be developed and encouraged, it also creates vulnerabilities when these experts retire or leave.

Scientific challenges

These commercial challenges cause or are exacerbated by some key scientific challenges.

Speed and resilience – perhaps the most obvious consequence of the commercial challenges discussed above is that R&D organisations need teams that can respond to market pressures and changing circumstances with effective innovation, and that can do so *quickly*. 'Doing what we have always done' is rarely enough. On the other hand, teams that are racing to solve the latest problem may not have much scope for re-engineering their whole methodology. Thus, tools or approaches that can speed up innovation and support the iterative development of new approaches while also fitting into existing workflows are of high value.

The cost of (repeated) experiment – a key factor slowing materials development is its high reliance on time-consuming testing. This applies throughout the materials development value chain, from characterisation of new materials, through qualification of materials and the parts they are used to make during product design, to quality assurance in manufacturing. But much of this testing could be unnecessary – either because it repeats measurements done in other projects or development phases, or because it relies on experimental designs that don't explore design space in the most efficient way. Methods that draw on existing data to increase experimental efficiency could be worth $millions and years of precious development time to materials R&D organisations.

> Methods that increase experimental efficiency could be worth $millions

Missing data – another limiting factor is the frequent sparsity of available data. It can cost tens of thousands of dollars to run the tests to measure the full set of properties that could characterise a material [9]. This means businesses often only measure the properties that they need for a specific project, rendering the data much less useful for future projects. And even material suppliers frequently provide only a subset of data in their product datasheets.

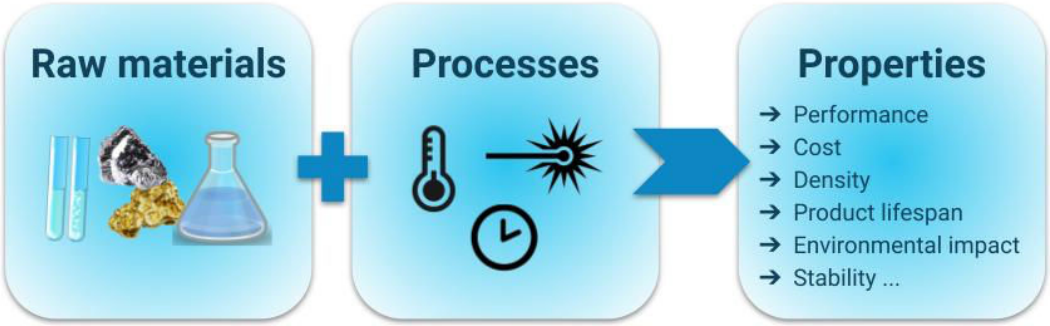


Figure 1. The 'formulation problem'

A high-dimensional, non-linear problem – whatever the material being developed, the fundamental problem is similar. We might call it the 'formulation problem' (Figure 1). We have a set of ingredients or raw materials and can vary the proportions and methods in which these are combined. Then we process the resulting material, for example, with mechanical, chemical, or thermal treatments. There are usually multiple processing steps. The result is a set of properties, and the challenge is to optimise these by varying the ingredients and processing. A basic grasp of mathematics tells us that even a few dozen alternatives for the ingredients, processes, and sequences in which they are combined quickly generates millions of possible combinations. And we may find that particular ingredients and processes, or the interactions between them, change properties in unexpected, non-linear ways. Material development problems can be daunting, and scientists need systematic approaches that can narrow down their options and identify the factors with the greatest impact on outcomes.

Capturing knowledge – Finally, R&D teams want to address these challenges in ways that not only deliver effective solutions, fast, but are also repeatable and lessen dependence on intellectual property that might be lost when people leave or retire. Can they capture knowledge as they go along, informing and speeding-up future projects?

How machine learning can help

The potential

Machine learning (ML) methods have the potential to address many of these challenges. In general terms, ML is a class of Artificial Intelligence (AI) that uses existing data on the inputs and outputs of a system to train a mathematical model that can then be applied to study the system. Figure 2 shows the relevance to materials development. The 'raw materials' here could refer to: the metals and alloying elements for an **alloy**; the resin, additives, fibres, and

fillers in a **plastic** or **composite**; or the inorganic material and additives in a **ceramic**. Examples of 'processes' include mechanical methods such as extrusion or moulding, heat treatments or curing, surface treatments, and layup of composites into layered structures. The properties of interest might be physical, mechanical, economic, environmental, or even aesthetic.

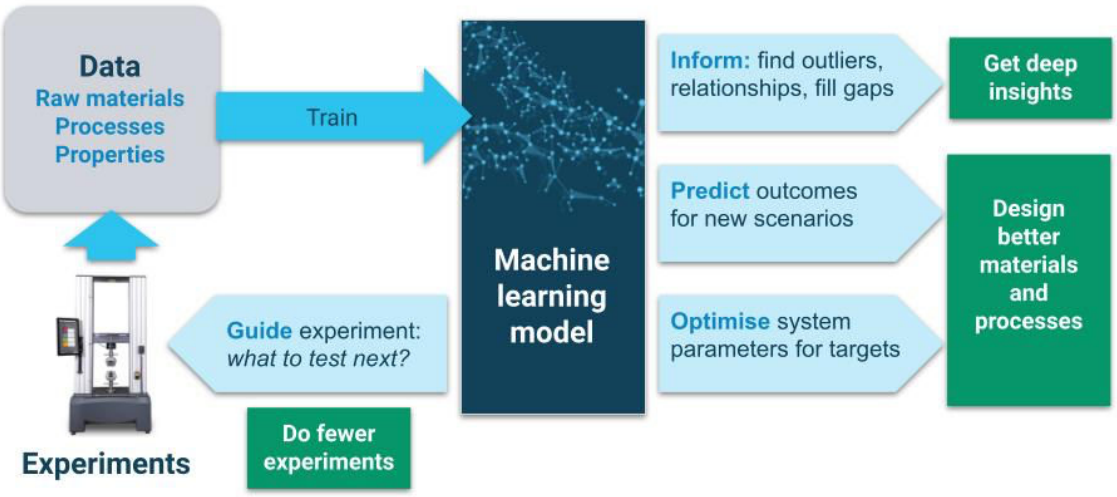


Figure 2. Machine learning applied to materials.

There are many different ML methodologies, but all involve the idea of learning from a set of 'training data' that includes inputs and the measured outputs in which those inputs result. The ML method will start from some generic representation, or set of representations, of the system and adjust the parameters in these representations and how they are combined until the output result is as close as possible to the measured outputs. The model generated in this way could be compared to an equation (albeit, often an enormously complicated one) that can then be used to predict outputs for a new set of inputs. The more training data that is fed into the model, the more accurate it should become.

As figure 2 shows, this provides the potential for some very interesting capabilities from a materials design and development point-of-view:

1. **Deep insights** – by varying inputs to the model and seeing how these impact on the outputs, it becomes possible to very quickly plot and understand key relationships within your materials system. For example, you can see which raw materials or process parameters are likely to have the greatest impact on your properties of interest, and how variations in these inputs might change the outputs.

2. **Designing better materials and processes** – by predicting the likely outcome for a new set of inputs ('virtual experimentation') or asking the model to find the set of inputs most likely to deliver a set of target outputs ('optimisation') you can identify improved material designs or optimise process parameters.

3. **Reducing experimental workloads** – the goal of any materials study is a model that consistently predicts performance within desired tolerances and boundaries. In conventional experimental projects, this means doing enough experiments to find and characterise the critical relationships within the system. Statistical Design of Experiments (DOE) approaches are used to cover design space as efficiently as possible, increasing the chances of adequately characterising the material. But they still result in relatively high experimental workloads. In an ML approach, not only is the identification of critical relationships already built into the modelling, but it is also possible to analyse the model as an experimental programme progresses to understand what missing data is most likely to improve model performance. This information can be used to guide what experiment to do next. Such adaptive experimental design [10] has been shown to result in 50-80% fewer experiments when compared to use of conventional DOE.

> Adaptive design of experiments can reduce experimental workloads by 50-80%

It is also worth noting the potential of the ML model to **capture knowledge** about a materials system in a re-usable way. A model can be re-used or refined for similar projects, providing information about the likely relationships in the system that could otherwise be lost.

Examples of ML delivering each of these benefits in real industry projects are shared below. Before discussing these application examples, however, we should review why use of ML is not yet more widespread and commonplace, and how this is changing.

Difficulties with applying ML

Three fundamental difficulties can constrain the use of ML in materials science.

> Three issues constrain use of ML: data limitations, fit to workflows, and managing trust and uncertainty

Data limitations – as we saw above, materials data is often sparse and noisy. And ML methods often fail when trained with such data. These methods may require clean datasets, where all values are complete for all features being studied. At best, this means that significant massaging of data is usually needed before ML methods can be applied to material problems.

Fit to workflows – there is a usability issue with much ML. Methods can be hard to apply and interpret. Use often requires coding or scripting or familiarity with ML concepts. Translating the requirements of a materials R&D project ("What should I test next?", "Which process parameter has the greatest effect on the strength?") into the language of an ML study ("What is the R-squared of my model?", "How should I set my hyperparameters?") can be confusing.

Trust and uncertainty – the fact that the ML model is a complex network, rather than an easily-inspected equation, makes it powerful when capturing those critical relationships, but means it cannot be easily understood without the right 'explainable AI' tools. This means that R&D

teams may hesitate to trust it. We also discussed above how ML models improve with more training data. The flipside of this is that their predictions will vary in accuracy, and, in fact, a given model may be more accurate for some parts of design space than for others. This is not a problem if you can quantify the accuracy of any given prediction, so that you can select options not simply on how well they seem to achieve objectives, but also on their likelihood of success. Coupling ML to accurate uncertainty quantification is thus important.

Introducing Alchemite™

The method

Intellegens has developed the Alchemite™ machine learning software to overcome these difficulties. The software is based on the Alchemite™ method [11] originally developed at the University of Cambridge. This proprietary algorithm, based on multiple imputation ML methods, is proven for training machine learning models **even when the training data is sparse and noisy**. This enables it to extract value from typical experimental and process data. Other key characteristics include:

- The ability to **fill in missing data**, enabling studies to assess materials for which property information is incomplete
- Accurate **uncertainty quantification** for predictions, enabling rational decision-making
- Guidance for the user on **what experiment to do next** (using Bayesian optimisation methods, in which the accurate uncertainty quantification plays a key role)
- A fast algorithm with a **light computational footprint**, generating results quickly.

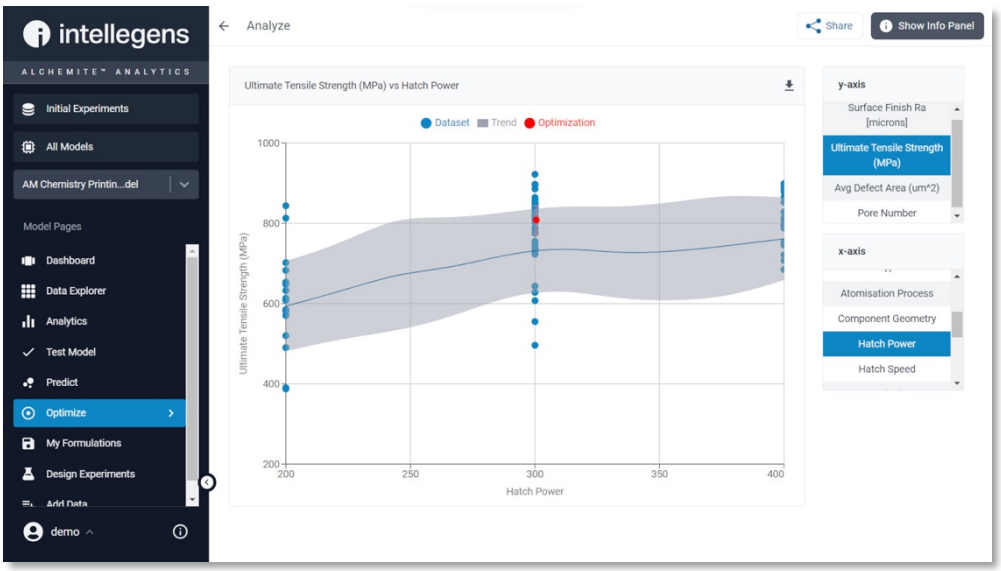


Figure 3. Alchemite™ Analytics web user interface: analysis of additive manufacturing data.

The software

The Alchemite™ software [12] is provided within a web user interface (figure 3), making it easy to deploy for use in materials R&D teams where machine learning and data science expertise is limited. A powerful API version also provides experienced data scientists with the option to integrate the Alchemite™ method into their own workflows and tools.

A key feature of the web user interface is the explainable AI tools – graphical analysis capabilities that make it easy to interpret the results and query the machine learning model. For example, the **importance chart** (figure 4) is a heat map that enables the user to understand what the model is telling them about which system inputs have the greatest impact on the system outputs.

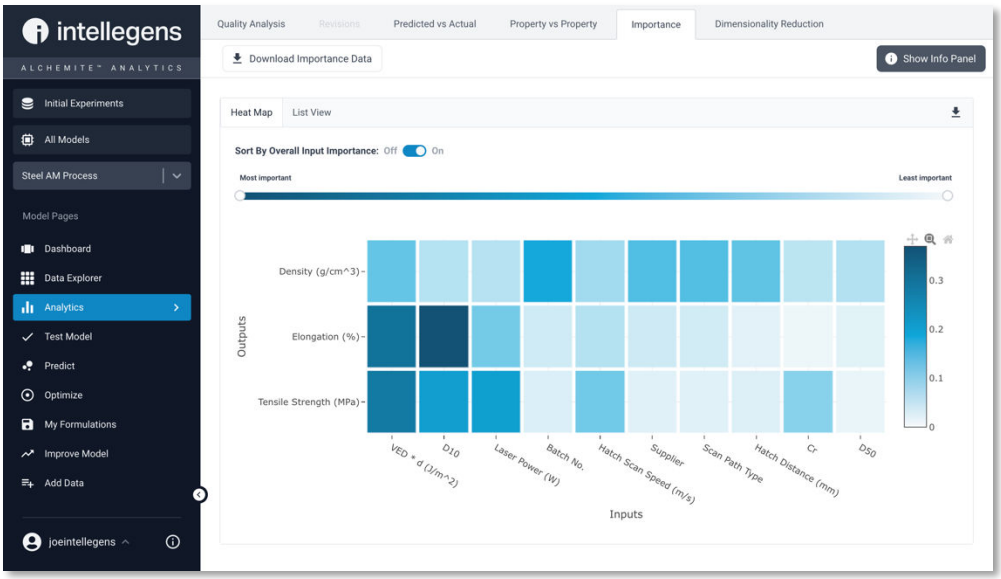


Figure 4. Importance chart; one of the explainable AI tools.

Application examples

These Alchemite™ features have enabled it to be applied successfully to a broad range of materials development challenges. Some examples are provided here. As we have published a separate white paper on polymer R&D [13], these examples focus on alloys and composites.

Gaining insight – Steel properties at OCAS (ArcelorMittal)

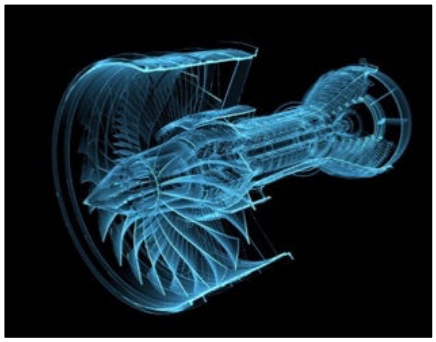
Alchemite™ was applied at **OCAS** (a joint venture between **ArcelorMittal** and the Flemish regional government) to understand steel performance [14]. An interesting feature of this project was that it included processing of microstructure images to generate numerical representing the images that was then used as an input to Alchemite™. This enabled the team to study the vital Process – Structure – Property relationships in steels.

Benefits of the project were assessed as:

> "The project has … provided us with insights that can help improve steel properties and focus valuable experimental resources."
>
> *Lode Duprez*
> *Chief scientific officer, OCAS*

- Alchemite™ deep learning allows extraction of hidden information from microstructural images.
- Alchemite™ provides models with good predictive power, going from processing and /or microstructure to properties.
- The model becomes less of a 'black box' when supported by analytics features, such as importance charts and sensitivity plots, and through providing reliable uncertainty estimates on predictions.
- The ability to deal with sparse and noisy data is essential for extracting useful information out of real-world datasets.

Materials design – Aerospace alloys at Rolls-Royce



The Alchemite™ technology was first validated in a collaboration between **Rolls-Royce** Aerospace and the University of Cambridge on the computational design of a new nickel-base superalloy for aero engine applications. The project found the material most likely to simultaneously fulfil eleven different physical criteria. This work was published in the journal *Materials Design* [15]. Key outcomes:

- The project team designed a new alloy to satisfy 11 physical criteria that was experimentally validated as outperforming existing alloys.
- The design process involved 90% fewer costly experiments.
- Such a reduction has potential to reduce costs by $millions and cut years from discovery and validation times.

Optimising properties – Hardfacing materials at Welding Alloys Group

Wear is one of the most challenging problems faced by heavy industry. In mining alone, 17% of the consumed energy is used to combat wear failure, accounting for 2.7% of global CO2 emissions. One of the most common methods to combat wear is by welding highly alloyed consumables (hardfacing materials) onto the surfaces of components. These must meet stringent requirements on safety, cost, environmental impact, and performance. Wear performance is determined by a complex interaction of properties, therefore optimising cost/benefit for hardfacing materials is difficult.

> "Technical know-how and data provided by WAG plus the ML algorithms provided by Intellegens resulted in a new composition that deviates considerably from materials used in the past."
>
> *Mario Cordero*
> *Group innovation director (WAG)*

Welding Alloys Group (WAG) and Intellegens applied Alchemite™ machine learning to this problem [16], which resulted in the development of an improved hardfacing material with dramatic cost/benefit advantages, not only from a performance, but also from an environmental point of view. This material has been taken to industrial-scale performance trials (fig 5).

Figure 5. An industrial component treated with the new hardfacing alloy (image courtesy of Welding Alloys Group).

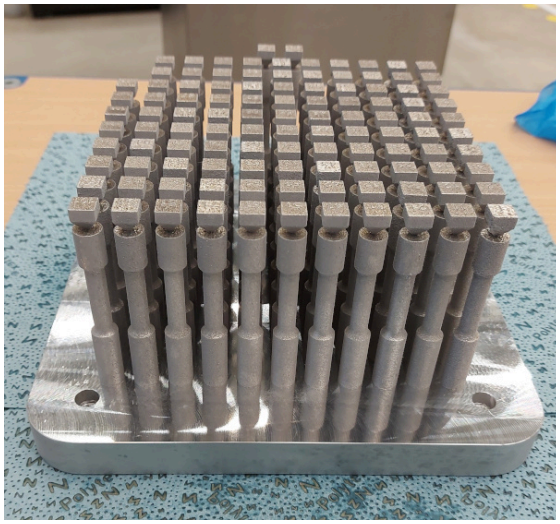


Figure 6. Additively manufactured test samples built during phase 2 of Project Medal.

Optimising processes and capturing knowledge – Additive Manufacturing with Project MEDAL

> "Machine learning has the potential to be a key technology in accelerating development and adoption of AM."
>
> *Lukas Jiranek*
> *Boeing*

ML can make the additive manufacturing (AM) process for metallic alloys in aerospace cheaper and faster. In collaboration with the University of Sheffield **Advanced Manufacturing Research Centre** (AMRC) and **Boeing**, Intellegens applied Alchemite™ to design new AM parameter sets for laser powder bed fusion (LPBF) and test them for nickel base alloys across two experimental cycles. Results from the second cycle (samples pictured in fig 6) showed good agreement with the predictions. The model was then used to develop process parameters for the new additive-specific Aheadd® CP1 powder from **Constellium** in combination with the 400W M2 machine from **GE Additive**. The project team was able to move from the new powder to final parameters in just

two builds, while applying no expert statistical knowledge. This shows how Alchemite™ machine learning can dramatically speed up AM process parameter optimisation. It also demonstrates the role of ML in capturing vital knowledge, since the model developed in this project can be applied to similar projects in future, speeding up identification of optimal process parameters for other new materials.

Design of Experiments – Composites at AMRC

Laminated fibre-reinforced polymer matrix composites are widespread in aerospace and increasingly used in other sectors, such as automotive. Identifying optimal cutting parameters for a given tool-composite pair can considerably reduce component non-conformance. Surface delamination during machining has been reported to result in 60% of all part rejections during final aircraft assembly. Understanding and controlling delamination requires significant experimental effort, typically involving systematic drilling and analysis of a thousand holes.

> "Alchemite™ could enable an 80% reduction in the direct costs associated with testing."
>
> *AMRC case study report*

Alchemite™ was used to reduce this experimental time by quantifying complicated nonlinear tool-composite relationships. It enabled analysis, understanding, and sharing of complex data relationships, guiding tooling design and selection prior to experimental campaigns. In this study at the **AMRC** Alchemite™ was able to deliver useful predictions of future tooling performance from sparse and noisy data based on 80+% fewer experiments and identified irrelevant features for predicting tool performance, facilitating further experimental cost savings.

Conclusion

Machine learning can accelerate innovation in materials R&D, but only if it can overcome some key difficulties – notably, the needs to deal with missing data, to ensure that it can be applied and interpreted by domain experts, and to provide confidence in its results through effective explainable AI tools and accurate treatment of uncertainty.

By meeting these challenges, the Alchemite™ software has provided valuable insights into materials properties, enabled the design and optimisation of materials and processes, and has delivered reductions of 50-80% in experimental workloads. This enables materials R&D teams to accelerate innovation while addressing challenges including volatile supply chains and costs, meeting tough environmental objectives, regulatory constraints, and the need to capture project knowledge more effectively.

References

[1] ‘Metal Global Market Report 2023’, Report: The Business Research Company, Media, PA, USA, 2023

[2] ‘Plastics Market Size, Share & Trends Analysis Report by Product, by Application, by Region, and Segment Forecast’, Report: Grand View Research, San Francisco, CA, 2022

[3] ‘Composites Market by Fiber Type, Resin Type, Manufacturing Process, End-Use Industry and Region – Global Forecast to 2027’, Report: Markets and Markets, Northbrook, IL, USA, 2023

[4] ‘Vehicle lightweighting’, Report: Automotive World, Penarth, UK, 2022

[5] M. Rossi, ‘The race to produce green steel’, *Ars Technica*, 19 June 2022. Available online: https://arstechnica.com/science/2022/06/the-race-to-produce-green-steel/

[6] X. Shen, L. Myllyvirta, ‘China’s steel sector invests $100bn in coal-based steel plants, despite low profitability, overcapacity, and carbon commitments’, Report: Centre for Research on Energy and Clean Air, 1 August 2023. Available online: https://energyandcleanair.org/publication/chinas-steel-sector-invests-usd-100-bn-in-coal-based-steel-plants-despite-low-profitability-overcapacity-and-carbon-commitments/

[7] L. Juliano, J.P. Bender, R. Russell, D. Hughes, ‘Gauging the Risks of Raw Material Volatility’, Report: BCG, Boston, MA, USA, 2022. Available online: https://www.bcg.com/publications/2022/gauging-risks-of-raw-material-price-volatility

[8] ‘The REACH Regulation’, The European Commission, 2023. Available online: https://environment.ec.europa.eu/topics/chemicals/reach-regulation_en

[9] Sepe M., ‘Getting the most out of your plastics data’, Prospector® Knowledge Center, 2014 Available online: https://knowledge.ulprospector.com/932/pe-getting-the-most-out-of-your-plastics-data/

[10] ‘Machine learning for adaptive experimental design’, Intellegens white paper, 2023

[11] ‘Alchemite™ - enabling applied machine learning’, Intellegens white paper, 2023

[12] Alchemite™ Analytics software, Intellegens, Cambridge, UK – https://Intellegens.com

[13] ‘Machine learning for polymer R&D’, Intellegens white paper, 2022

[14] ‘Improving steel performance with OCAS (ArcelorMittal)’, Case study, 2021. Available online: https://intellegens.com/understanding-and-improving-steel-performance-with-ocas/

[15] B.D. Conduit, N.G. Jones, H.J. Stone & G.J. Conduit, 'Design of a nickel-base superalloy using a neural network', *Materials & Design* **131**, 358, 2017

[16] 'Improving hardfacing consumables with Welding Alloys Group', Case study, 2020. Available online: https://intellegens.com/machine-learning-achieves-drastic-improvement-in-hardfacing-welding-consumable/

[17] 'Data-driven Additive Manufacturing: AMRC, Boeing, Constellium, GE Additive, 2021. Available online: https://intellegens.com/data-driven-additive-manufacturing-with-amrc-and-boeing/

[18] 'Composite manufacturing with the AMRC', 2022. Available online: https://intellegens.com/composite-manufacturing-with-the-amrc/

About Intellegens

Our mission is to be the leading machine learning solution for real-world, sparse and noisy data problems in industrial R&D and manufacturing processes. Our focus is on making it easy to apply machine learning to accelerate innovation. Alchemite™ originated at the University of Cambridge and development is on-going at Intellegens, in close collaboration with our growing community of Alchemite™ customer organisations. These represent sectors including alloys, additive manufacturing, aerospace, batteries, ceramics, chemical processes, composites, consumer products, cosmetics, drug discovery, energy, food and beverage, formulated products, paints, plastics, and printing technology.

www.intellegens.com | info@intellegens.com | @intellegensai