WHITE PAPER

Machine learning for polymer R&D

Accelerating innovation for plastics, elastomers, paints, and other polymeric products

intellegens.com | info@intellegens.com
Intellegens, Eagle Lab, 28 Chesterton Road, Cambridge, CB4 3AZ, UK

intellegens

Applied machine learning

Executive Summary

Polymer R&D supports applications in almost every industrial sector. There is a continued need for fast innovation, not only to find performance improvements, but also to address challenges including cost and energy price volatility, sustainability, regulatory constraints on ingredients and processes, and supply chain disruption. Achieving speed and resilience requires teams to understand complex, high-dimensional, non-linear relationships in polymer formulations in a context where key 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 polymer formulation space, to predict and test potential solutions, and to focus experimental work. But machine learning methods can struggle where data is sparse, as it often is in polymer R&D. And these methods can also be difficult for polymer domain experts to apply and interpret. We introduce **Alchemite™**, machine learning software that overcomes these difficulties. We discuss an application example in which Alchemite™ has been used to reduce experimental workloads and provide valuable insight into polymer blending. Similar reported applications include lubricants, paints, inks, and materials design.

The needs of polymer R&D

Commercial challenges

Challenges include cost and energy price volatility, sustainability, regulatory constraints, and supply chain disruption

Consumer products, packaging, tyres, structural materials, paints, coatings, medical devices, clothing, sports goods, household goods, toys – polymers are ubiquitous in everyday products. The global plastics industry was worth nearly $600 bn in 2021 [1]. The proliferation of new materials since the development of the first synthetic plastic, Bakelite, in 1907 has been spectacular. Today, one well-known database records over 100,000 commercial plastics and elastomers [2]. But polymer R&D is far from done. The process of blending polymers and additives to optimise performance offers almost infinite formulation possibilities. In exploring this space, R&D teams must meet commercial challenges including:

- **Cost and energy** – cost pressures are always present, exacerbated by current volatility in global energy prices which means that the economics and competitiveness of process routes are subject to constant change.

- **Environmental** – polymers are deeply entwined with the oil-based economy that has driven the climate crisis. The impact of single-use plastics on the marine environment is now widely recognised. Product innovation alone cannot solve these challenges, which require major systems changes and government intervention. But innovation can help, for example, by increasing recycled content, more effective use of biopolymers, and controlling negative impacts as plastics degrade.

- **Regulatory** – constraints are increasingly imposed on polymer R&D by regulations designed to protect both the environment and human health. Additives or fillers may suddenly disappear from the supply chain or have limitations placed on their use. R&D teams need to respond quickly.

- **Supply chain issues** – global supply chains are increasingly sensitive to disruption, which can be caused by all the factors outlined above, and more. Can R&D teams make their products and manufacturing processes more resilient?

- **Intellectual property** – perhaps more than any other area of materials science or chemistry, polymer science is frequently viewed as an 'art'. This is partly because, at the molecular level, polymers consist of long chain molecules, often tangled in arrangements that are much less ordered than, for example, those in metals or ceramics. Understanding and precise control of properties can be difficult, and subtle changes in how polymers are blended, processed, and the additives used can make big differences. This leads to a high dependence on the intrinsic knowledge of experts and thus to problems for businesses when these experts retire or leave.




Scientific challenges

How do these commercial challenges translate into scientific challenges for R&D teams?

- **Speed and resilience** – above all, R&D organisations need teams that can respond to emerging cost, environmental, regulatory, or supply chain challenges with effective innovation, and that can do so fast. Tools and approaches that speed up innovation are of high value.

- **Missing data** – one limiting factor is the sparsity of available data. It can cost tens of thousands of dollars to run all of the tests required to measure the full set of properties that typically characterise a polymer grade [3]. 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 polymer suppliers frequently provide only a subset of data in their product datasheets.

Polymer R&D has a high reliance on costly, time-consuming experimentation

- **A high-dimensional, non-linear problem** – polymer properties are typically achieved by blending different polymer or resin materials together with fillers and additives. Processing steps also impact properties – both those used during formulation, such as heat treatments, and final manufacturing processes, such as extrusion or moulding. As we have already discussed, all of this, together with the 'organic' nature of polymer materials, makes understanding and controlling polymer properties an 'art'. In mathematical terms, the system is high-dimensional and non-linear.

- **High reliance on costly experimental testing** – in turn, this complexity leads to a high reliance on empirical approaches in polymer science, supported by large amounts of testing. As we already saw, such tests are costly. They are also time-consuming.

- **Capturing knowledge** – polymer R&D teams want to address these challenges in ways that not only deliver effective solutions, quickly, but are also repeatable. Can they capture knowledge as they go along, informing and speeding-up future projects and reducing their exposure to experts leaving the organisation?

Can machine learning help?

The potential

Machine learning methods have the potential to address many of these challenges. In general terms, machine learning 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. Using the model, we can explore the impact on the system's outputs of changing its inputs, predict the outputs for a new set of inputs, or find optimal inputs to meet a target set of outputs. We can improve the model by adding more data to it. It is even possible to use the machine learning model to tell us which data to gather to achieve the best improvement for the least effort.

Figure 1 shows how these general capabilities apply when the system in question is a formulation – a chemical system with ingredients and processing parameters as its inputs and properties as its outputs. Polymers are often delivered as formulations or blends in which the ingredients are one or more polymers plus components such as fibres (to improve mechanical properties), additives (e.g., colourants or fire retardants), or other fillers (to alter performance or simply to reduce cost).

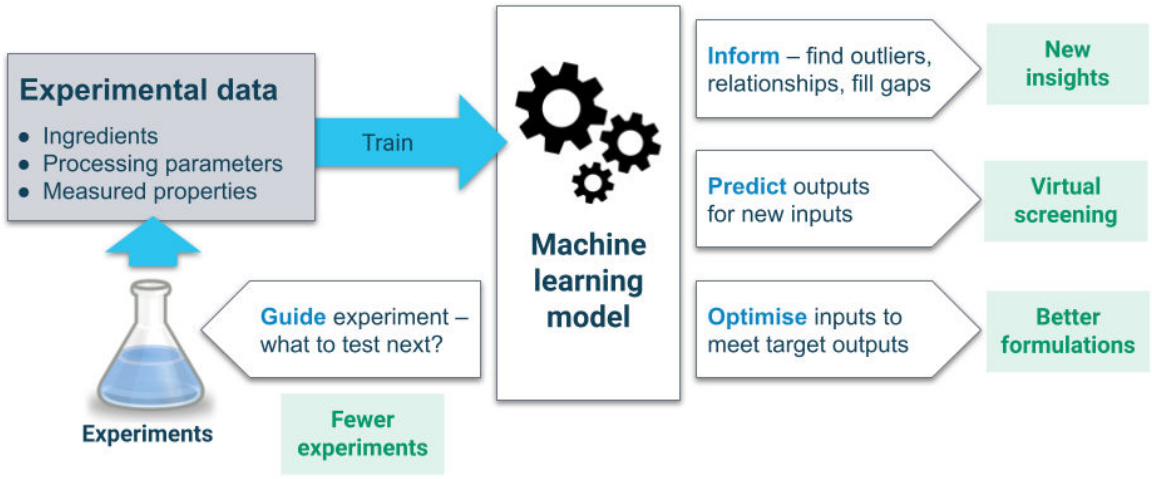

Figure 1. Machine learning applied to formulations.

In this case, the model is built using experimental data and allows us to gain insights about how ingredients and processing influence properties, to screen potential new formulations virtually, and to design optimal formulations. By guiding us on what to test next, the model enables an adaptive design of experiments approach [4] which can be highly efficient in maximising understanding of the formulation while reducing the amount of experiment required – typical reductions range from 50-80%.

So, machine learning can address the scientific challenges outlined in the previous section. Models can capture high-dimensional, complex, non-linear relationships. We can gain valuable insight and identify potential solutions, faster, while reducing experiment. And models can be re-used predictively, capturing valuable knowledge.

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

Difficulties

There are, however, some problems with applying machine learning in polymer R&D. A key difficulty comes from the issue of missing data. Polymer data is frequently sparse. And machine learning methods often fail when trained with sparse data – they require clean datasets where all the values are completed for all the features being studied. At best, this means that significant massaging of data is usually needed before machine learning methods can be applied to polymer problems.

There is also a usability issue with much machine learning. Methods can be hard to apply and to interpret. Using software often requires coding or scripting skills or familiarity with machine learning concepts. The fact that the underlying mathematical model is a complex neural network, rather than an equation that can be easily inspected, makes it powerful when capturing those complex relationships, but also means it cannot be easily understood without the right explainable AI tools. These factors contribute to making it difficult for the polymer scientists and chemists in a typical polymer R&D team to benefit from machine learning.

Introducing Alchemite™

The method

Intellegens has developed the Alchemite™ machine learning software to overcome these difficulties. The software is based on the Alchemite™ method [5] originally developed at the University of Cambridge. This proprietary algorithm 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 characteristics of the method include:

- The ability to **fill in missing data**, enabling studies to assess polymers 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)
- A fast algorithm with a **light computational footprint**, generating results quickly.

The software

The Alchemite™ software [6] is provided within a web user interface (figure 2), making it easy to deploy for use in polymer 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.

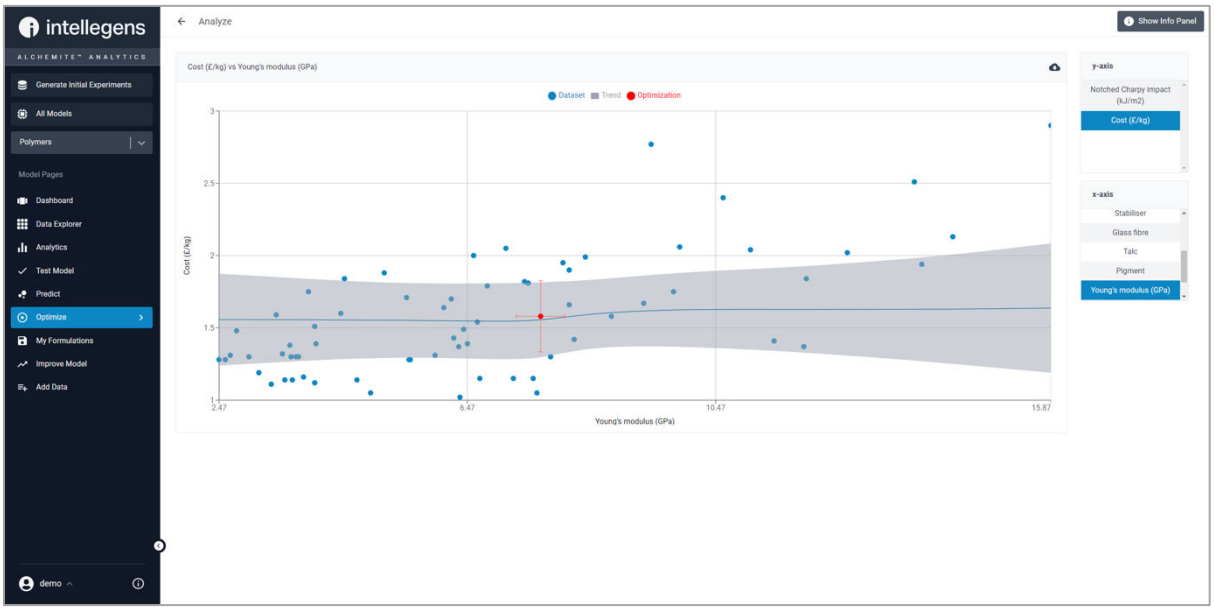

Figure 2. Alchemite™ Analytics web user interface showing analysis of polymer data.

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 3) 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.

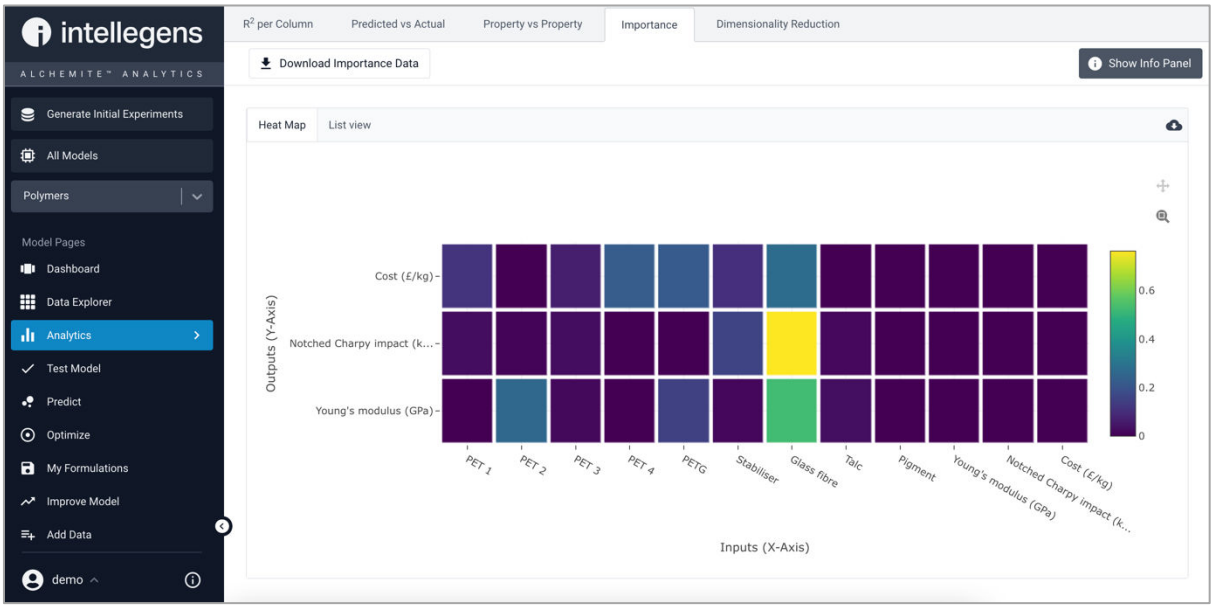

Figure 3 – Importance chart; one of the explainable AI tools.

Application example

Problem

The Alchemite™ software has been successfully applied to many projects in polymer R&D. In this example project, based on Intellegens work with plastics producers, Alchemite™ was used to **guide an experimental program studying a polymer blend**. The aim was to learn how to vary each type of monomer or constituent ingredient in the polymer while also optimising process conditions, to improve the performance of the blend. An important constraint was controlling cost. The existing experimental data set was sparse and consisted of relatively few tests since each new blend was expensive to make. The initial priority was to enrich this data by doing the right experiments to develop a model that was predictive of the blend properties while reducing experimental cost when compared to the use of traditional methods.

Solution

Alchemite™ was used to build a model from the existing experimental data. As part of the process, the Intellegens Science Team “featurised” the data, using the chemical properties of each monomer. This process involves representing distinct chemical features of the monomer in the data so that the machine learning can learn from them and relate these

features to outputs, building up an understanding of the similarities and differences between monomers.

Once a model was built, the team reviewed its quality, measured via the Alchemite™ Quality Metric that shows how accurate the model's predictions are when applied to a test set of data. Alchemite™ was then used to suggest which ten blends should be made next to gain the most improvement in the performance of the model for the least effort. Once made, the properties of these blends were measured, and this data was fed back into Alchemite™ which then re-trained its model. The new model could be assessed for quality and analysed to understand the polymer blend system.

Outcomes

The new experiments generated an improvement in model accuracy substantially greater than any likely improvement from experiments chosen using standard Design of Experiments methods. Thus, understanding of the polymer blend was improved with fewer experiments than would normally have been the case – saving time and cost. The model could be inspected using explainable AI tools, for example, to understand how changes in the ingredients affected the cost and properties. This understanding informs recipe changes that enable the blend to achieve the same performance but at lower cost.

Understanding of the polymer blend was improved while saving experimental time and cost

The project has also created a predictive model that can be re-used. One potential application for this is responding to changes in the composition of the input ingredients once the polymer is in production. This can happen when using recycled polymer as an input, since there can often be significant variability in the composition of recycled feedstock over time. By applying the model, it will be possible to rapidly propose changes to the process parameters and other ingredients to keep the properties of the polymer within specification.

Other applications

Other examples of applying Alchemite™ to polymers or closely related systems have included:

- Predicting the physical properties of **lubricants**, in a collaboration with **BP** [7]
- Combining data from simulation and experiment to optimise **design of materials** including polymers at **Lucideon** [8]
- **Formulation of paints** at a major global paint producer.
- **Design of experiments** for the formulation of inks at **Domino Printing Sciences** [9]
- **Filling in missing data** in a polymer property database with **Ansys Granta** [10].

Conclusion

Machine learning can accelerate innovation in polymer R&D, but only if it can overcome some key difficulties – notably, the needs to deal with missing data and to ensure that it can be applied and interpreted by polymer domain experts. By meeting these challenges, the Alchemite™ software has provided valuable insights into problems such as polymer blending, while substantially reducing the amount of costly and time-consuming experimentation required to achieve project goals. This enables polymer R&D teams to accelerate innovation while addressing challenges including cost, energy consumption, environmental objectives, regulatory constraints, and the need to capture project knowledge more effectively.

References

[1] "Plastic Market Size, Share & Trends Analysis", Report: Grand View Research, San Francisco, CA, 2022

[2] UL Prospector® database, UL Solutions, Northbrook, IL.

[3] Sepe M., "Getting the most out of your plastics data", Prospector® Knowledge Center, 2014 [Online] https://knowledge.ulprospector.com/932/pe-getting-the-most-out-of-your-plastics-data/

[4] "Machine learning for adaptive experimental design", Intellegens White Paper, 2021

[5] "Alchemite™ deep learning – solving complex problems with real world data", Intellegens White Paper, 2021

[6] Alchemite™ Product Information [Online] https://intellegens.com/products-services/

[7] Santak P., Conduit G., "Predicting physical properties of alkanes with neural networks", *Fluid Phase Equilibria*, 501, 112259, 2019. https://doi.org/10.1016/j.fluid.2019.112259

[8] "Transforming materials and process development through machine learning", Intellegens Webinar, 2021 [Online] https://intellegens.com/webinar-transforming-materials-and-process-development-through-machine-learning/

[9] "Formulation design with Domino Printing", Intellegens Case Study, 2020 [Online] https://intellegens.com/machine-learning-for-virtual-experiments-ink-formulation/

[10] Verpoort P.C., MacDonald P., Conduit G.J., "Materials data validation and imputation with an artificial neural network", *Computational Materials Science*, 147, 176-185, 2018. https://doi.org/10.1016/j.commatsci.2018.02.002

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