

Surrey Uni on challenging AI decisions

12 February 2026



AI systems already decide how ambulances are routed, how supply chains operate and how autonomous drones plan their missions. Yet when those systems make a risky or counter-intuitive choice, humans are often expected to accept it without challenge, warns a new study from the University of Surrey.

Epsom and Ewell Times adds that the Civil Aviation Authority has granted Amazon a licence to deliver items by drone. It is uncertain when this service will actually begin.

The research, published in the Annals of Operations Research, looked at the use of optimisation algorithms in relevant areas such as transport, logistics, healthcare and autonomous systems. Optimisation algorithms are systems that decide the best possible action by weighing trade-offs under fixed rules such as time, cost or capacity. Unlike prediction models that estimate what will happen, optimisation algorithms choose what should be done.

Optimisation algorithms decide what gets prioritised, delayed or excluded under strict limits such as weight, cost, time and capacity. Yet those decisions are mathematically correct but practically opaque.

The research team's findings implies that our increasing 'blind trust' creates serious safety and accountability risks in the increasing areas of everyday life where optimisation algorithms are used.

Using a classic optimisation challenge known as the Knapsack problem, the research demonstrates how machine learning models can learn the structure of an optimisation decision and then explain it in plain language. The method shows which constraints mattered most, why certain options were selected and what trade-offs pushed others out.

The study shows how organisations can challenge optimisation algorithms before their decisions are put into practice. Rather than replacing existing systems, the approach works alongside them, using machine learning to analyse decisions and explainable AI to reveal why one option was chosen over another and which constraints and trade-offs shaped the outcome.

Dr Wolfgang Garn, author of the study and Associate Professor of Analytics at the University of Surrey, said:

"People are increasingly asked to trust optimisation systems that quietly shape major decisions. When something looks wrong, they often have no way to challenge it. Our work opens those decisions up so humans can see the logic, question it and intervene before real-world consequences occur."

This is particularly important for autonomous systems such as delivery drones. Drones must constantly decide which packages to carry while balancing battery life, payload weight and safety requirements. Without transparency, regulators and operators cannot easily justify or audit those decisions.

Rather than replacing existing optimisation software, the approach works alongside it. Machine learning is used in this approach to analyse solutions, explain feasibility and identify brittle or high-risk decisions before deployment.

The research introduces a structured framework that ensures explanations are tailored to real decision makers. Instead of technical outputs, systems can provide human-readable reasoning, such as: "too many heavy items were selected, or battery limits were prioritised over delivery value."

Dr Garn continued:

"Regulators are starting to ask harder questions about automated decisions. If you can't explain why your system chose one option over another, you'll struggle to get approval — or defend yourself when something goes wrong. This framework makes that explanation possible."

Surrey University



Photo credit www.routexl.com. Llicence <https://creativecommons.org/licenses/by/2.0/>