January 2026

Towards Self-Healing Supply Chains

An Exploration in Safeguarded AI

Written by

HASH

Funded by

Advanced Research + Invention Agency | ARIA

Towards Self-healing Supply Chains

An Exploration in Safeguarded AI

Table of Contents

Introduction **3**
Methodology **5**
About the authors **5**
Part 1: The Domain **6**
Overview 6
Biopharmaceuticals 6
Safety critical 8
Part 2: Challenges **9**
Current responses 10
Part 3: AI adoption and sentiment **12**
Ambition and vision 12
Planning and strategic coherence 12
Implementation drivers and organisational risk 12
Data governance, security, and deployment constraints 13
Effective operating models for AI adoption 13
Human oversight and trust 13
Manual adjustment 14
Part 4: Safeguarding **15**
The gatekeeper concept 15
Part 5: Use cases **16**
Predictive Temperature Management 16
Manufacturing Golden Batch 18
Proactive Supplier Management 20
Production Optimisation 22
Long-range Capacity Planning 24
Supply Chain Disruption Response 26
Part 6: Path to deployment **28**
Deployment Challenges 30
Overcoming Challenges 30
Part 7: Next steps **32**

Introduction

Supply chains are the circulatory system of our global society, with goods, services, and parts crossing international borders multiple times as part of production and distribution.

But they are vulnerable and often fragile, and increasingly complex and globally distributed. Widespread adoption of just-in-time operating models reduce costs but leave little buffer in the form of inventory or spare capacity.

Supply chain disruption is a key risk to the global economy, alongside climate change and geopolitical instability[1]. Pandemics, conflict, extreme weather, and transportation issues have repeatedly caused cascading effects. Even local disruptions can quickly propagate across entire networks, with outsized economic and operational consequences.

Biopharmaceutical supply chains are particularly important and fragile. Errors are costly and frequently irreversible, directly affecting patient outcomes. New medicines have introduced patient-specific material flows and data and coordination demands too complex and time-critical for human management alone.

Making supply chains more resilient and efficient therefore has far-reaching economic and societal benefits. Robust supply chains reduce the likelihood and severity of disruptions, helping to stabilise production, protect economic output, and support sustained growth. Greater efficiency lowers operating and logistics costs, which in turn helps keep prices down for businesses, governments, and consumers.

In safety-critical sectors such as healthcare and biopharmaceuticals, improved resilience and coordination can have even higher stakes: ensuring the timely availability of medicines, vaccines, and therapies, reducing shortages and saving lives.

The increasing power of Artificial Intelligence (AI) provides a key opportunity to transform and dramatically improve how supply chains are managed, with significant potential value to be gained[2].

But it is not without risks. Many current AI systems are sensitive to poor or changing data, can produce confident but incorrect outputs, and often lack sufficient explainability for regulatory scrutiny. In the biopharmaceutical industry, AI-driven decisions can directly affect product quality, regulatory compliance, and patient safety, raising the bar even further for reliability and trust.

Without robust safeguards, AI risks introducing new failure modes into already safety-critical environments, making safety assurances a prerequisite for widespread adoption.

---

[1] https://www.weforum.org/publications/global-risks-report-2023/

[2] https://www.mckinsey.com/capabilities/operations/our-insights/supply-chain-40-the-next-generation-digital-supply-chain

As a result of these challenges, the UK's ARIA is funding the exploration of applications for "Safeguarded AI"[3] – AI controllers with mathematical safety guarantees – which have the potential for transformative impact.

This report describes research conducted over 2025 which explored the key challenges facing supply chain management, and the potential for and barriers to the deployment of Safeguarded AI solutions.

Throughout, there is a particular focus on biopharmaceutical supply chains, but many of the findings and opportunities have wider relevance.

It will be of interest to people:

1. who work in supply chain management (SCM) and want to find out about industry trends, attitudes towards AI, and emerging technology to address key challenges
2. who work in AI/AI safety and want a deeper understanding of applications and challenges in supply chain management
3. with a general interest in either area.

[3] https://www.aria.org.uk/opportunity-spaces/mathematics-for-safe-ai/safeguarded-ai/

Methodology

The work carried out over 2025 included:

- Conducting more than a hundred in-depth interviews with 80+ biopharmaceutical SCM professionals across 40+ companies to capture operational detail, decision-making context, and critical safety criteria.
- Building a library of supply chain processes and translating them into mathematical models in collaboration with the University of Oxford and other ARIA-funded creators.
- Working with industry stakeholders to identify pilot AI solutions to validate the feasibility, safety, and potential impact of the most promising applications.
- The creation of representative synthetic supply chain data to underpin modelling, experimentation, and solution development.

About the authors

HASH is a venture-backed startup building a next-generation AI platform focused on knowledge integration, decision optimization, and automation for complex, data-rich environments.

The HASH platform emphasises safety and auditability, making it suitable for contexts where reliability and trust are critical.

HASH is working with ARIA to pilot use cases for trusted AI in supply chain management, as well as developing the software that allows users to define, review and monitor Safeguarded AI solutions.

For more information, or if you have any questions regarding this report, please contact:

- Dei Vilkinsons <d@hash.ai> - *SAILS Co-lead*
- Ciaran Morinan <c@hash.ai> - *SAILS Co-lead*

Part 1: The Domain

Overview

Supply chain management in the broadest sense is concerned with the movement of products, money & information.

It aims to match supply of products and services over time to demand, and decide what to do in the event of mismatches between planned supply and forecast demand, while optimising for cost & efficiency.

Supply chains contain a number of nodes – manufacturing locations, distribution centers, third-party suppliers, customers – interconnected in a complex non-linear network, with many thousands of different types of items moving between them.

Human decision-making in this context can be slow, rely on insufficient information, and ultimately be unable to fully consider and envisage the complex ripple effects of any decision, particularly in the event of sudden and large-scale disruption.

Sub-optimal and delayed decisions can result in large financial impacts. Over a decade, a single prolonged shock to production in a pharmaceutical supply chain is estimated to cost the equivalent of 24% of one year's EBITDA[4]. Over the COVID-19 period (2020-21), world trade would have been ~2.7% higher cumulatively without supply chain disruptions[5].

Supply chain management represents the ultimate multi-parameter optimisation problem, with many problems across industry being able to be framed in supply chain terms (e.g. allocating limited resources to fulfil a need). The potential value to be gained through the use of AI in supply chain management is significant, in reducing forecast errors, lost sales, and administrative costs[6].

Biopharmaceuticals

Biopharmaceutical supply chains are even more complex than those in most other industries because they sit at the intersection of regulation, biology, intellectual property, and patient safety.

Pharmaceutical products span several fundamentally different categories, each with distinct scientific, manufacturing, and supply chain characteristics:

- **Small-molecule pharmaceuticals:** chemically synthesised, typically stable, and relatively easy to manufacture at scale (e.g. paracetamol). They often have longer shelf lives, can tolerate a wider range of temperatures, and are suited to

[4] https://www.mckinsey.com/capabilities/operations/our-insights/risk-resilience-and-rebalancing-in-global-value-chains
[5] https://www.ecb.europa.eu/press/economic-bulletin/focus/2022/html/ecb.ebbox202108_01~e8ceebe51f.en.html
[6] https://www.mckinsey.com/capabilities/operations/our-insights/supply-chain-40-the-next-generation-digital-supply-chain

high-volume, centralized production. Supply chains tend to be global, cost-optimised, and comparatively simpler and more flexible.

- **Biologics**: large, complex molecules produced using living systems such as cell cultures (e.g. insulin, vaccines, monoclonal antibodies). Manufacturing is far more sensitive to process conditions, with inherent variability in yields and longer, less predictable cycle times. Typically require cold-chain handling, have shorter shelf lives, and are difficult to scale or transfer between sites. Supply chains are therefore more capacity-constrained, risk-sensitive, and tightly coupled to specific manufacturing assets and regulatory approvals.

- **Advanced therapies**: therapies often involve living cells, viral vectors, or patient-specific materials (e.g. cell and gene therapies). Manufacturing is complex, failure rates can be high, and products are extremely sensitive to time, temperature, and handling. Supply chains are often regional or local, with tightly choreographed logistics linking collection, manufacturing, testing, and administration, with resilience and traceability critical.

- **Personalised and patient-specific therapies**: manufactured from an individual patient's own cells (e.g. autologous cell therapies). The supply chain becomes a closed-loop system connecting a single patient with a clinic and manufacturing site. Forecasting, inventory buffering, and traditional planning approaches largely break down. Identity preservation, chain of custody, and chain of identity become central design requirements rather than supporting controls.

The pharmaceutical industry is increasingly moving toward biologics and personalised medicines due to their ability to target disease mechanisms more precisely and address previously untreatable conditions (thus, most companies are classed as *bio*pharmaceutical).

These new modalities are expected to generate 60% of revenue in the next five years, significantly increasing manufacturing and supply chain complexity[7]. The industry is currently evolving to run their existing small-molecule supply chains as efficiently as possible, while developing new and innovative ways to supply the new product classes.

They must do so while dealing with several specific challenges:

- **Regulatory fragmentation and compliance**: With multiple, overlapping regulatory regimes across regions and countries. Differences in regulatory approvals, labeling rules, packaging requirements, and market authorizations mean that materials and finished products often cannot be shared, substituted, or reallocated across markets, even when products are chemically identical.

- **Quality control:** Raw materials, intermediates, and finished products are subject to extensive testing, release procedures, deviation management, and formal investigations. A minor process deviation or documentation error can lead to batch rejection or prolonged quarantine. Unlike many industries, quality

[7] https://www.bcg.com/publications/2025/emerging-new-drug-modalities

decisions directly block or enable physical flow, creating tight coupling between quality systems and supply chain operations.

- **Shelf life constraints and expiry risk**: Pharmaceutical products have strict, finite shelf lives from manufacture. Many customers require a minimum remaining shelf life (e.g., six months). Stability data, storage conditions, and regulations limit inventory holding time and usage. This heightened sensitivity to demand uncertainty and disruptions makes inventory planning challenging and increases costs for excess or slow-moving stock.

- **Temperature management and cold-chain dependence**: Many products need controlled temperatures during manufacturing, storage, and distribution. Temperature deviations can cause investigations or total write-offs, even if the product is undamaged. Maintaining cold-chain integrity adds significant monitoring, validation, and risk-management complexity.

- **Long lead times and limited recoverability**: End-to-end lead times, from raw material sourcing through manufacturing, testing, release, and distribution, are often long and rigid. Once a disruption occurs, recovery options are limited by regulatory and technical constraints. There is little ability to expedite, substitute, or rapidly increase capacity compared to other industries, and delays at earlier stages often compound downstream as planned manufacturing or transportation slots are missed.

Safety critical

Biopharmaceutical supply chains already operate close to their safety limits. They are also full of edge cases: rare disruptions, conflicting constraints, incomplete data, and regulatory nuance that is not fully captured in historical examples.

Human planners manage these systems through conservative heuristics, buffers, and manual controls. As biologics, cell therapies, and personalised medicines become more prevalent, the system becomes too complex, time-critical, and data-rich for manual coordination alone.

Errors can directly harm patients, compromise public trust, or violate regulatory obligations.

Because these systems affect patient safety and public health, “mostly safe” behaviour is not sufficient. The industry requires AI systems that are safe by design, not just safe on average.

In the coming sections of this report we explore live challenges facing the industry, the state of current AI adoption, and the potential for ‘guaranteed safe’ AI solutions.

Part 2: Challenges

This section draws on interviews and pilot engagements with organisations across manufacturing, logistics, and supply chain management. Participants included large multinational firms, specialist logistics providers, and technology startups.

Interviews covered end-to-end supply chain activities, including network design, demand and supply planning, product security and traceability, supply resiliency, logistics, cold-chain management, transportation, and manufacturing operations.

Across stakeholders, the challenges described were consistent and systemic, reflecting structural issues rather than isolated operational inefficiencies:

- **Manual processes and low-value work:** Many supply chains remain highly manual. Significant expert time is consumed by low-value activities such as compiling data for presentations, creating quality tickets, and maintaining planning parameters. This diverts scarce expertise away from analysis, risk mitigation, and continuous improvement.

- **Informal, slow and intuition-driven decision-making:** Decision logic is often held mentally and negotiated informally, rather than formally represented, stress-tested, or consistently applied. Decisions often need to be made quickly, particularly during disruptions, yet the information required is slow to assemble.

- **Complexity and hidden trade-offs:** Decisions involve many interacting variables across the supply network, making second- and third-order impacts difficult to anticipate. Organisations seek to optimise while maintaining safety and regulatory compliance, but these trade-offs are hard to quantify.

- **Operational risk and regulatory complexity:** Small operational errors can have disproportionate consequences, including stock-outs, excess inventory, or patient impact, amplified by high product value and limited shelf life. Product security threats such as diversion and counterfeiting, along with fragmented global regulations, further increase complexity and operational burden.

- **External supply and disruption risk:** Disruption risk is difficult to detect early, particularly in external supply chains managed by third parties. Limited visibility into supplier operations constrains proactive intervention and risk mitigation.

- **Logistics and distribution:** Import and customs processes introduce significant uncertainty, with clearance times varying widely by country. Transport disruptions and temperature excursions are common and require rapid response, often triggering investigations, delays, or write-offs.

- **Manufacturing and batch release:** Manufacturing often suffers from unplanned downtime, long changeovers, and frequent plan changes, lowering capacity and increasing risk. Fragmented process knowledge complicates root cause analysis, and a reliance on specialists creates bottlenecks.

Current responses

Historically, biopharmaceutical companies have approached these challenges by prioritising control, standardisation, and compliance over automation or intelligence. The dominant strategy has been to impose structure on inherently complex operations through enterprise systems, a myriad of formal processes, and layered governance, rather than attempting to model or reason about the system end to end.

- **Large transactional systems, and specialised planning, quality, manufacturing and logistics platforms were introduced to centralise data, enforce workflows, and create auditable records.** These systems reduced some variability and enabled global scale, but they were not designed to eliminate expert effort or support rapid reasoning. As a result, human specialists remain responsible for stitching together information across systems, reconciling inconsistencies, and translating system outputs into decisions. Spreadsheets, macros, and ad-hoc scripts are the de facto glue, compensating for system rigidity and gaps while embedding critical logic outside formal tools.

- **Decision-making is formalised primarily through governance processes rather than computational representation.** Regular meetings, escalation pathways, and cross-functional reviews are used to negotiate trade-offs and resolve conflicts. Analytical support comes via deterministic planning models, heuristics, and scenario tools that explore predefined alternatives but are slow to adapt and fragile. Much of the true decision logic – how trade-offs were weighed, how risks were interpreted, when rules could be bent – remains tacit, in experts' heads rather than in systems that can be stress-tested or reused.

- **Where analytics and AI are applied, they tend to be narrow, localised, and retrospective.** Statistical process control, basic forecasting models, optimisation routines, and rule-based alerts help identify deviations, optimise individual parameters, or highlight exceptions after thresholds are crossed. These tools improve visibility and consistency but rarely support anticipation or holistic reasoning across manufacturing, supply, quality, logistics, and regulatory constraints. Complexity is managed by decomposition: each function optimising its own objectives, with integration handled manually through meetings and negotiation.

- **Risk management follows a similar pattern. Instead of predictive or preventative intelligence, organisations rely on buffers, redundancy, and procedural controls.** Safety stock, lead-time padding, conservative product release rules, and extensive documentation absorb uncertainty but at high cost and with slow response times. External supply and logistics risks are monitored through lagging indicators, infrequent and time-consuming audits, and manual reviews, limiting the ability to detect weak signals or intervene early. Compliance requirements further reinforce sequential, document-heavy processes that prioritise traceability over speed or adaptability.

These approaches have succeeded in making biopharma supply chains auditable, compliant, and scalable, but not truly adaptive. They reduce some forms of error while entrenching manual effort, implicit reasoning, and slow response.

Rudimentary AI and analytics acted as decision aids rather than decision partners, supporting experts without capturing or operationalising their judgement. This legacy explains why expert time remains scarce and overloaded, and why more advanced approaches are now being sought to formalise decision logic, reason under uncertainty, and safely reduce human bottlenecks.

As a result of the continuing challenges, the industry is increasingly exploring potential transformational value of more advanced AI solutions.

Part 3: AI adoption and sentiment

In contrast to the broad consensus on challenges, the appetite for, approach to, and concerns about the use of AI varied widely across those we spoke to.

Ambition and vision

Some organisations are working towards highly autonomous supply chains capable of sensing disruptions, selecting corrective actions, and executing responses with minimal human intervention. Others are limited to cautiously experimenting with generative AI for limited productivity improvements, such as drafting meeting minutes or summarising documents.

Ambition does not correlate strongly with company size. Both large multinationals and smaller, niche providers, particularly specialist providers like those in cold-chain logistics, were amongst those leading in experimentation and deployment.

Despite this variation, a common aspiration emerged across almost all respondents; the desire to shift human effort away from reactive, administrative firefighting toward proactive, strategic decision-making.

Planning and strategic coherence

Only a small number of companies (typically those with the most ambitious visions) had developed clear, staged roadmaps for AI adoption.

Even when there was a clearly articulated overarching vision, most AI use cases were conceived and executed in isolation. Interviewees frequently described frustration with this siloed approach, noting that it limited cumulative impact and made it difficult to scale successful pilots beyond local use.

Implementation drivers and organisational risk

In some organisations, adoption is driven bottom-up, fuelled by grassroots enthusiasm and opportunistic experimentation by motivated individuals or teams. In others, projects are initiated top-down primarily because “AI is the topic of the moment,” rather than as a response to a clearly defined problem or value case.

Both patterns carry risk. Grassroots initiatives often struggle to secure sustained sponsorship or integration into core systems, while hype-driven projects can consume resources without delivering meaningful or durable benefits. Without a unifying strategy, organisations risk accumulating disconnected tools that increase complexity rather than reducing it.

Data governance, security, and deployment constraints

At larger firms in particular, interviewees highlighted strong hesitancy around adopting external AI solutions, often due to strict internal rules governing data usage, cybersecurity, and procurement.

Some tools operate entirely within an organisation's existing infrastructure or on-device, while others require sensitive data to be transmitted to external cloud environments. The latter often trigger lengthy cybersecurity reviews and compliance checks, significantly slowing experimentation and deployment.

Differences between providers in how data is handled are not always transparent, making it difficult for practitioners to assess which solutions are viable. Even when vendors do not train models on customer data, the mere transmission of sensitive information outside organisational boundaries can pose significant compliance challenges.

Effective operating models for AI adoption

The most effective users of AI were those who combined strong in-house domain expertise with external specialist support, most often from product-led startups rather than traditional consultancies. This reflects both the rapid pace of change in AI capabilities and the need for continuous iteration and experimentation.

AI services are rarely "finished" products at this stage of the technology lifecycle. New approaches that are cheaper, faster, or more reliable emerge within months or even weeks. Organisations are therefore increasingly partnering with startups that can continuously incorporate these advances, rather than investing heavily in long-running consultancy engagements.

Human oversight and trust

Across many interviews, full automation was not seen as desirable in the near term. Instead, respondents expressed strong interest in systems that allow humans to verify, approve, and modify AI-generated decisions before they take effect. This human-in-the-loop approach lowers adoption barriers and allows AI outputs to be aligned with domain expertise and contextual knowledge.

However, these capabilities are rarely well supported in practice. AI-generated outputs are often difficult or impossible to edit, limiting their usefulness and reducing trust.

Many practitioners also cited a lack of visibility into the reasoning behind AI recommendations as a major barrier to adoption, particularly in safety-critical situations where patient outcomes could be affected.

Even when AI systems demonstrate low error rates, the absence of traceability, explainability, and assurance often leads users to replicate the underlying analysis manually to verify a recommendation. In such cases, AI fails to reduce workload and instead adds an additional layer of process.

Manual adjustment

Practitioners using existing planning tools routinely rely on manual adjustments and informal reasoning to compensate for known model blind spots, whether or not AI is involved.

Common examples included systematic forecast bias in certain markets, historical “black swan” events such as COVID distorting forward-looking estimates, planning parameters that do not reflect real-world operations, and anticipated regulatory or trade changes such as tariffs that are not yet encoded in systems.

This further underscores the need for AI systems that can incorporate uncertainty, be safely evolved and adjusted by human experts, and be evaluated on their impact on safety and optimisation for any state of the world, rather than tested against a fixed list of historical scenarios for which the correct answer is already known.

The Safeguarded AI approach is designed to offer this resilience, flexibility and assurance.

Part 4: Safeguarding

The gatekeeper concept

Most current AI systems optimise performance metrics without formal models or guarantees. They are trained on historical data, tested on held-out datasets, and assumed to generalise safely. This assumption breaks down in biopharmaceutical supply chains, where rare events matter most and the cost of failure is high.

There is no reliable way to enumerate all unsafe decisions in advance, nor to test every possible scenario. Post-hoc monitoring or "human-in-the-loop" checks can help, but are insufficient if AI systems are to act at speed or scale.

The Safeguarded AI programme treats safety as a first-class engineering problem, using mathematical modelling techniques to evaluate the impact of AI decisions against safety criteria and optimisation goals.

It relies on several components, chiefly:

1. **A world model:** a mathematical representation of the space in which the AI is operating, sufficiently detailed to realistically simulate the downstream consequences of any action it takes or recommends, including accounting for areas of uncertainty. In a supply chain context, this might include the inputs and outputs of network nodes and processes, fluctuations in demand, variability in manufacturing and distribution times, and quality control failures.

2. **A safety specification**: a formal description of what outcomes are acceptable and unacceptable, with reference to observable states of the world. In biopharmaceuticals this would encode patient safety (e.g. on-time availability of critical medicines), regulatory compliance (e.g. process conformance), quality thresholds, security constraints, and environmental limits, as well as business outcomes (e.g. reducing cost and waste, and improving delivery time).

3. **A verifier:** a mechanism that produces an auditable proof that an AI controller trained using the Safeguarded AI method will satisfy the safety specification relative to the world model.

Given a world model and safety specification, the AI training process can produce a machine-checkable proof certificate attesting the probability of compliance with safety criteria under any number of scenarios – historical, imagined, and generated – without the need for the 'right' answer to be known in advance. More details can be found on the programme's website.

By embedding formal safety guarantees into AI systems, the industry can unlock their transformative potential without unacceptable risk.

Part 5: Use cases

The Safeguarded AI approach is predicated on an AI taking or recommending an action, the consequences of which can be modelled to assess whether they are safe and desirable.

Ideal use cases therefore involve systems which suggest specific actions, rather than prediction, monitoring or classification systems which indirectly influence decisions.

But given the relative immaturity of AI adoption in the domain, and barriers to and unmet preconditions for the rollout of decision-making systems, in practice the road to Safeguarded AI is likely to go via solutions playing a more supporting role.

In Part 6 we discuss the path to deployment in general terms, including challenges.

Here we describe a selection of concrete use cases in the biopharmaceutical industry and in supply chains more broadly. More can be found on our website.

Proactive Supplier Management

Problem

Quality and performance management of external suppliers is traditionally fragmented, retrospective, and labour-intensive. In biopharmaceutical supply chains, a large share of production is outsourced to vendors and contract manufacturing organisations (CMOs), yet oversight typically relies on periodic audits, manual document reviews, static scorecards, and post-hoc deviation investigations.

Signals of declining performance - such as subtle increases in deviations, delayed deliveries or responses, documentation quality drift, or regulatory scrutiny - are often not detected or investigated until more significant and persistent issues have emerged. This reactive posture increases the likelihood of supply disruptions, prolonged investigations, regulatory exposure, and in the most severe cases patient impact.

Solution

An AI-enabled quality oversight system continuously integrates internal and external quality-relevant data to build a dynamic, contextual view of vendor performance and proactively identifies, reports on and ultimately remediates performance risks across the external supplier network.

The system assesses the likelihood and extent of disruption or quality issues and recommends or takes action to remediate any impact on operations or customers. Actions may include working with the existing supplier (from requests for information to more intensive audits), shifting orders to established alternative suppliers, or exploring new suppliers.

Modelling

This use case involves predicting outcomes such as cost and time of order fulfilment, quality issues, and operational disruption, including simulating the impact of:

1. projected performance of current suppliers, under different possible scenarios (e.g. what is the outcome if performance worsens or improves at various rates).
2. remedial action taken with the current supplier (e.g. how much does it cost, how likely is it to improve performance, at what rate).
3. alternative sourcing (e.g. additional cost and time needed to switch volume to existing alternative or new suppliers).

The system aims to ensure stability, quality and cost-efficiency of supply.

Data

- **Internal quality and manufacturing data**: quality deviations, Corrective and Preventive Actions (CAPAs), change controls, batch disposition outcomes, right-first-time metrics, investigation cycle times, audit findings, and quality events linked to specific vendors, sites, products, and batches.

- **Enterprise and supply data**: data on materials, products, and suppliers; production volumes; demand criticality; inventory buffers; and dependency relationships between products, sites, and suppliers.

- **Vendor-supplied information**: Quality documentation, responses to observations, periodic reports, certificates, and correspondence that may signal changes in responsiveness, maturity, or compliance posture.

- **External and regulatory signals**: Public inspection outcomes, warning letters, recalls, enforcement actions, and other regulatory intelligence associated with vendors, sites, or peer manufacturers.

Much of this information exists today in siloed, partially structured, or unstructured forms, making it difficult to synthesise and analyse without automated support. The first step on the path to Safeguarded AI would involve ingesting the data into a holistic view of supplier performance, which already offers value in providing early warning signals, even before safeguarded AI recommendations for actions are layered on top.

Safety

The safety criteria for this system would include that:

1. quality issues are not left unaddressed.
2. recommended changes in suppliers do not lead to expected interruptions in critical supply (e.g. due to ramp-up/onboarding time).

Impact

Proactive supplier management moves quality oversight from retrospective audits to real-time assurance, delivering significant risk reduction. 19% of organisations estimate their financial exposure to a major third-party incident at US$500 million or more; 11% at more than US$1 billion.[8]

On top of this, Safeguarded AI decisions for corrective action further drive faster, better responses to issues, with effective simulation of potential outcomes.

Production Optimisation

Problem

Optimisation in complex manufacturing networks is traditionally constrained by static assumptions embedded in planning systems that fail to reflect the true performance of assets, processes, and logistics flows.

Production schedules are often built on nominal cycle times, yields, and transport lead times that diverge significantly from reality. Over long order cycles, this mismatch accumulates into bottlenecks, idle capacity, excess inventory, missed delivery commitments, and unnecessary expediting.

In make-to-order environments with high demand volatility and capacity constraints, these gaps create both material delivery risk and missed opportunities to increase throughput and customer service.

There can often also be multiple options for configuring production processes, including choice of manufacturing sites and raw material sourcing, with the impact of decisions harder to assess as complexity grows.

Solution

An AI system recommends the parameters that drive planning outcomes, such as production rates, changeover assumptions, yield factors, and transportation lead times.

It takes account of variability in timings and outcomes, and competing demands on production resources, in order to optimise throughput. In more complex versions it does so across a large network, taking into account optionality in what resources are used as well as the parameters producing the plan.

During live operation, it continuously monitors plan adherence and responds to discrepancies with analysis of the downstream impact on operations and customer commitments, and provides recommendations for change.

---

[8] https://www.deloitte.com/uk/en/services/consulting-risk/research/the-imperative-for-increased-supply-chain-resilience.html

Modelling

The production process is represented in a formal model which includes:

1. Products, including the production process associated with them, as well as their relative importance to the business and to customers (to inform trade-offs when balancing production across multiple products).
2. The network of nodes involved in production processes, including relationships between them, transportation time, etc.
3. The probability distribution of expected time to complete each process step.
4. Accounting for expected downtime of assets (e.g. for maintenance or changeover), as well as scenarios covering occasional unplanned downtime.

Data

- **Enterprise resource planning systems**: Planned and actual production events, raw material orders and receipts, quality inspections, customer orders and deliveries, and inter-site transfers.
- **Advanced planning systems**: Future production and transport plans and associated parameters.
- **Manufacturing execution systems**: Machine runtimes, batch consumption and production data, and changeover performance.

Safety

Recommendations are evaluated against predefined operational and safety constraints, such as workforce limits, equipment runtime boundaries, and plan risk tolerances, ensuring that optimisation never violates real-world safety and feasibility.

When balancing across multiple products, recommendations are evaluated against their optimisation of the overall portfolio (e.g. the system should not optimise one production process by worsening outcomes for another, more important process).

Impact

AI-driven production optimisation enables a shift from reactive firefighting and underutilised assets toward a more predictive, resilient, and economically efficient manufacturing network. “Smart manufacturing” which takes full advantage of digital technologies can unlock 10-20% more throughput from existing assets[9] and a 30-50% reduction in total machine downtime[10].

---

[9] https://www.deloitte.com/dk/en/blogs/cxo-board/blog-tore-s-mart-manufacturing-2025-Technologys-rise-and-why-it-still-comes-down-to-people.html
[10] https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/digital-in-industry-from-buzzword-to-value-creation

Predictive Temperature Management

Problem

Cold chain logistics management traditionally relies heavily on passive monitoring and reactive interventions. Temperature-sensitive pharmaceutical products are shipped with one or more temperature monitors, with a ‘temperature excursion’ occurring when readings report a breach of the acceptable range (e.g. 2–8°C cold chain, 15–20°C ambient, -15°C / -50°C frozen).

Any excursion prompts an assessment of whether the product is still fit for purpose, which weighs the extent and duration of its exposure to suboptimal conditions against its known durability under them.

Temperature and humidity is typically tracked through manual inspections or basic sensor logs, leading to lagging visibility and slow corrective responses to excursions. This leads to increased risk of delays, wastage and regulatory non-compliance.

Solution

A Safeguarded AI system would continuously ingest and analyse real-time sensor & contextual data to forecast whether a shipment is trending toward a temperature excursion *before* limits are breached, and triggers timely action to intervene and prevent or remediate the excursion.

Such interventions could include sending the product via a different route, or pausing transit to stabilise the temperature prior to ongoing movement.

In addition, where intervention is not possible or is unsuccessful in preventing an excursion, an AI-driven product degradation model could use all historical storage and transit temperature data to determine whether or not the product is still suitable for use up to the printed expiration date.

Modelling

The world model for this solution would include:

1. Actual historical and predicted future environmental conditions for each stage in the product’s journey.
2. Available options for intervention in the event of an excursion (for example, alternative shipping routes, with associated assumptions about journey time and environmental conditions).
3. The product’s assumed tolerances to and degradation under different environmental conditions.
4. The availability of alternative supply of the product (in the event that it must be replaced to fulfil an order).

The consequences of AI decisions would be simulated using this world model to ensure that, in the event of an expected or predicted temperature excursion, products are routed according to business and customer priorities while ensuring patient safety.

Safety

The system must:

1. never recommend actions that increase excursion risk.
2. never allow products to be delivered which have failed or require an assessment following a temperature excursion.
3. always correctly detect and respond to temperature excursions.
4. when choosing to reallocate or redirect products, ensure that overall business and patient needs are met (e.g. don't reallocate from a higher priority destination to a lower one).

Data

- **Real-time telemetry:** temperature readings, GPS / location trace, device lifecycle events (start/stop, download, calibration)
- **Product and quality constraints:** allowable ranges (low/high thresholds), packaging configuration and qualification data (shipper type, coolant, validated duration profiles), product impact assessment reference data (e.g. technical reports), batch/lot identity and history (to support cumulative impact modelling)
- **Shipment/route and execution data:** route definitions and planned milestones, carrier / mode (air/sea/road/rail), lane identifiers, historical lead times and variability by lane/mode/season, transportation management events (handoffs, customs status, delays), ERP references (order, customer, priority), proof-of-delivery signals
- **External context:** Weather by geolocation and time (temperature, humidity, wind, extreme events), traffic/port congestion signals if available and relevant to duration variability
- **Business impact and prioritisation inputs:** Product criticality (medical criticality, shelf-life constraints, alternative supply availability), shipment value / penalties, demand urgency / stock-out risk (dynamic over time)

Impact

Temperature excursions and the resulting product wastage have a significant economic impact. Considering only vaccines in Europe, an estimated €4 billion a year is lost to waste in the cold chain.[11] Globally, and looking beyond vaccines to other

[11] https://www.iqvia.com/-/media/iqvia/pdfs/emea/library/whitepaper/tip-of-the-iceberg-economic-and-environmental-impact-of-the-vaccine-cold-chain.pdf

biopharmaceutical products, IQVIA estimates that global losses due to temperature excursions across all products in the pharmaceutical industry due to temperature excursions exceed more than $35 billion annually.

Predictive cold chain management directly mitigates these losses and delivers value across safety, service, cost, and resilience. By automatically detecting and mitigating temperature excursions earlier enterprises can reduce waste, reduce operational burden of manual management, and improve order fulfilment.

Manufacturing Golden Batch

Problem

Manufacturing technology transfer, which scales pharmaceutical formulations from R&D labs into full-scale production, traditionally relies on iterative experimentation, manual parameter tuning, and the tacit knowledge of experienced operators.

Process recipes are adjusted step by step to account for differences in equipment size, geometry, and dynamics, with initial batches often carrying elevated risk of suboptimal yield, variability, or deviation.

Limited reuse of historical learning across products and equipment, combined with fragmented data and conservative safety constraints, leads to long development cycles, extensive physical trials, and avoidable cost and quality risk.

Solution

A Safeguarded AI system defines “Golden Batch” profiles by drawing on historical manufacturing data and equipment profiles to recommend amendments to processes.

The system learns from prior technical transfer attempts, recipe evolutions, and observed outcomes to recommend parameter bands that are statistically associated with higher yield, better quality, and more reliable execution. A built-in safety layer ensures that recommendations always remain within validated operational and quality limits, and rejects unsafe parameter combinations.

Modelling

The world model would include known properties of the manufacturing process, ingredients and equipment, and assumed outcomes under various conditions, in order to simulate the outcome for a given set of conditions.

This can include mechanistic models for important properties which affect process timings such as heat transfer rate and mixing time.

Data

Various sources of structured and unstructured data are required to enable this capability:

- **Process and recipe data**: R&D formulations, historical “Day Zero” and post-transfer manufacturing recipes, time-series evolution of recipe parameters, and version histories for individual process steps (e.g. mixing time, speed, temperature profiles).
- **Manufacturing execution and outcome data**: Batch records, in-process sensor data (actual temperatures, durations, speeds, pressures), deviations from planned recipes, yields, cycle times, quality outcomes, and any associated investigation findings.
- **Equipment and facility context:** Equipment type and configuration, vessel dimensions, rated operating ranges, safety thresholds, usage guidelines, and where available, estimated physical properties relevant to scale-up (e.g. heat transfer characteristics).
- **Quality and safety constraints**: Validated operational limits, critical process parameters (CPPs), critical quality attributes (CQAs), and any regulatory or site-specific rules that must never be violated by recommended parameter sets.
- **Tacit and external knowledge**: Operator insights, informal guidance captured from historical documents or interviews, mechanistic or semi-mechanistic models where available, and relevant external technical literature that informs process behaviour.

These data are integrated into a structured representation of manufacturing knowledge, capturing relationships such as which recipes were executed on which equipment, how recipes evolved over time, and what outcomes resulted under different conditions.

Safety

Safety conditions for manufacturing would include safe operating parameters and expected conditions for equipment, including for example:

1. Temperature
2. Heating rate
3. Pressure
4. Impeller rotation speed

AI recommendations for processes would be rejected if they violate (or are simulated as violating) these constraints.

Impact

Golden Batch methodologies drive an estimated 10% improvement in yield and a 30% reduction in batch rejections[12], and more advanced predictive methods can deliver value across development speed, quality, cost, and operational robustness.

They effectively reuse any volume of past data alongside predictive models to reduce reliance on repeated costly physical trials, improving first-time-right performance at scale, and enabling a more consistent, transparent, and trustworthy approach to manufacturing tech transfer.

Long-range Capacity Planning

Problem

Long-range capacity planning (usually spanning the next 5–10 years) is a critical but often underpowered capability in complex manufacturing and supply networks.

Decisions about building, expanding, consolidating, or retiring production assets are typically made using site-level analyses, static spreadsheets, and manually curated scenarios.

While these approaches can approximate future needs, they struggle to capture network-wide interactions, long lead times, regulatory constraints, and the compounding effects of demand uncertainty. As a result, capacity investments are frequently reactive, misaligned with true demand growth, or insufficiently resilient to future shocks.

Solution

At the core of the solution is an integrated model of the supply network that represents production, packaging, and warehousing capacities alongside multi-year demand scenarios. The model incorporates site-level attributes such as ramp-up times, flexibility, technical capabilities, and regulatory constraints, as well as strategic initiatives like portfolio changes, technology transfers, or policy shifts.

Rather than focusing on near-term scheduling, the solution simulates the interaction between demand and capacity over a 5-10 year horizon, providing strategic visibility into where and when future supply–demand mismatches are likely to emerge and recommending the optimal supply chain design to resolve these gaps.

Recommendations can include addition or removal of sites or manufacturing assets (whether internal or external), infrastructure investment, make-versus-buy strategy, site consolidation, and technology deployment.

---

[12] https://www.ey.com/en_us/insights/life-sciences/achieving-gold-standard-operations-in-pharma-manufacturing

Continuous monitoring of input assumptions can trigger alerts when conditions change, while proposed network designs can be simulated under different trajectories to stress-test against service, cost, and sustainability objectives.

Modelling

The world model takes a high-level view of over a long time horizon of:

1. Demand, e.g. forecasted by product type and region.
2. Supply, including the capabilities of the existing network and modifications to the network that would increase or decrease supply of different products.

A key benefit of the approach is in its ability to simulate the impact of changes under many thousands of capacity-demand scenarios, ensuring that decisions are tested under a wide range of multiple future states.

Data

Key data inputs include:

- Production and packaging capacities at SKU or campaign level from enterprise systems.
- Warehousing constraints from warehouse management systems.
- Long-term demand forecasts under baseline, upside, and downside scenarios, which can be expanded to many more during model training.
- Parameters describing site readiness, site capabilities, investment lead times, and operational flexibility.
- A range of data to support network changes, including estimates for building new facilities, consolidating sites, the cost of external suppliers, etc.

Safety

As the solution looks at balancing aggregate demand and supply, the key requirements are that:

1. There is not a significant shortfall in meeting demand (e.g. aggressive consolidation or decommissioning leading to an inability to meet demand for critical product)
2. Significant investment is made in increasing supply which is not required.

Given increasing uncertainty over time, it can only offer probability estimates for these requirements being met under different assumptions, but nonetheless provides a more robust and comprehensive analysis than other methods.

Impact

Long-range capacity planning shifts investment decisions from reactive to strategic. By creating a shared fact base for supply chain, finance, and strategy teams, the system ensures alignment on future network needs.

Optimising the network footprint helps organisations avoid unnecessary builds and supports resilience through diversification. Real-world case studies show a reduction of operating costs of up to 20% from network redesign[13,14].

Integrated network modelling also enhances agility. Gartner reports that connected planning accelerates strategic decision-making by 2-3x.

Supply Chain Disruption Response

Problem

Supply chain disruption management in pharmaceutical networks is traditionally a high-pressure, manual, and reactive process. When immediate imbalances between supply and demand arise teams must act quickly to prevent orders going unfulfilled.

Decisions such as whether to expedite shipments, reallocate inventory, increase production, switch sources, or do nothing and hope for the best are often made with fragmented, outdated data and limited ability to anticipate downstream consequences.

This results in suboptimal outcomes, unnecessary cost, avoidable emissions, operational disruption across other products, and elevated risk to patient supply.

Solution

A Safeguarded AI system would assess short-term supply–demand imbalances, evaluate feasible response options, and take or recommend actions that balance patient safety, service continuity, cost, and broader network stability. The objective is to respond to disruption without unnecessary or counterproductive interventions.

A critical feature of the approach is global rather than local optimisation. The system explicitly accounts for resource contention across shared inputs, production assets, labour, and logistics capacity. When evaluating an action to protect supply for one product or market, it also estimates the disruption cost imposed on others, preventing solutions that simply shift risk elsewhere in the network.

Modelling

The world model for responding to disruption includes:

1. Outstanding and expected customer orders.

[13] https://www.bcg.com/publications/2024/why-manufacturers-should-focus-on-cost-of-goods-sold
[14]https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/tech-enabled-business-transformation-the-trillion-dollar-opportunity/what-can-i-do/supply-chain-management

2. Available stock (including location and possibility of reallocation, taking account of market requirements).

3. Cost and timing for various modes of transportation.

4. Feasibility, cost and time for increased supply (e.g. through manufacturing)

The model is used to simulate the impact of different actions on downstream impacts across inventory availability, stock-out risk, service levels, cost-to-serve, and disruption to other products competing for shared resources.

Actions may include expediting shipment or switching transport mode to ensure time delivery, as well as non-standard actions that carry higher risk, such as shipping or producing more stock while desk-based quality reviews are ongoing, reallocating stock between markets, or reworking product.

Data

Key data inputs include:

- Inventory positions and demand profiles by product and location.
- Planned and actual shipment routes with observed lead-time distributions.
- Manufacturing schedules, changeover profiles, and capacity constraints.
- Bill-of-materials (BOM) and input material availability, including procurement, transport, and testing lead times.
- Quality status information such as quarantines and expected release timelines.
- Cost parameters covering freight, materials, overtime, obsolescence, storage, and contractual penalties.

Additional business rules and attributes such as priority customers, medical criticality by product and market, regulatory constraints, and environmental considerations are incorporated as explicit constraints and optimisation criteria.

Safety

Safety guardrails ensure that recommendations never violate regulatory requirements, approved routes, capacity limits, labour rules, or patient safety principles, and that disproportionate harm to specific geographies or patient groups is avoided.

Impact

The value of safeguarded supply chain disruption management lies in its ability to replace ad hoc, intuition-driven decision-making with a structured, explainable, and risk-aware process. By replacing manual intuition with AI-driven global optimisation, organisations can significantly reduce response times and improve overall outcomes.

Part 6: Path to deployment

The Safeguarded AI approach relies on a world model of processes in the domain and data to inform possible scenarios that might occur – sufficient to simulate and evaluate the impact of AI decisions under many possible conditions – as well as agreement on what constitutes safe and desirable outcomes.

The path to deployment for many companies will therefore include steps which lay those foundations, each of which delivers value in their own right, including:

1. **Data unification:** integrate data from planning, quality, manufacturing, logistics, and inventory systems and standardise descriptions of entities (products, sites, lanes, batches, vendors, orders) to provide a unified, comprehensive view.
2. **Document decisions and processes:** triggers, handoffs, constraints, approvals, and what "good" looks like (service, cost, compliance, patient criticality, emissions).
3. **Build monitoring and situational awareness:** deploy narrow, low-risk capabilities that reduce information lag and manual effort. These tools don't decide yet: they make humans faster, more consistent, and more evidence-driven, while generating training data on signals, actions, and outcomes.
4. **Add prediction and scenario generation:** introduce predictive models that estimate future risk states and trajectories.
5. **Introduce decision support with simulated consequences**: move from "what might happen?" to "what should we do?" – now the AI is recommending actions, guided by performance and safety requirements. Humans approve or reject, and the system logs decisions, reasoning and results to improve the knowledge base and world model.
6. **Introduce automated action:** once trust is established, move to full automation of suitable decisions. This will require:
    a. Well-defined, tested and formalised performance objectives, safety requirements and risk tolerances.
    b. A world model which has been informed by extensive historical data and validated by subject matter experts.
    c. Extensive training and stress testing of the system under a number of different possible scenarios.
    d. Guardrails that allow for escalation in critical or highly uncertain scenarios, and always human override and auditability.

Rollout can expand along three axes:

1. **decision criticality:** from less to more important decisions.
2. **network scope**: from a small slice (e.g. one product, site, or vendor) to broader portfolios and regions.
3. **autonomy:** from prediction and alerting systems only, through a small number of low-risk, bounded options which can be taken autonomously, to wide latitude in and options for action.

Deployment Challenges

Pilot implementations face several key challenges.

1. **Data availability:** accessing usable data is the most consistent barrier, particularly from heterogeneous manufacturing equipment where data may be local-only. Some use cases also require or at least benefit from new hardware (e.g. sensors) adding cost and friction. Even where systems exist, data quality, completeness, and security remain concerns.
2. **Tacit knowledge:** critical decisions, especially in planning and process control, rely heavily on practitioners' tacit knowledge, experience, and judgment, which are not formally captured in systems of record. This knowledge gap is acute for use cases like disruption response and supply allocation, where historical decisions and rationales are essential for AI training but sparse as data.
3. **External dependencies:** many high-impact supply chain challenges require collaboration with external partners (e.g. suppliers) who may be unable or unwilling to share data or provide system integrations due to contractual limits, technical immaturity, or commercial reasons.
4. **Data governance:** data that could reveal proprietary information requires careful handling, and potentially aggregation and anonymisation strategies, to comply with corporate policies while enabling external training and validation of AI systems.
5. **Focus:** ongoing large-scale system transformations (e.g. migrating to SAP S/4HANA) means that some organizations are not yet in a position to consider additional experimental pilots.

Overcoming Challenges

As part of our work to build a credible path to Safeguarded AI deployment with our industrial partners, we have been:

- Deliberately prioritising supply chain management pilots where decision logic is clearer, hardware dependencies are lower, and early value can be demonstrated with fewer prerequisites.

- Working intensively with subject matter experts in companies to understand and document existing processes, and translate tacit into explicit knowledge.
- Making the HASH core technology platform available free of charge to pilot partners, lowering barriers to engagement for organisations with low perceived AI readiness or data maturity. The platform supports integration of structured and unstructured data into a unified, strongly typed knowledge graph, creating a foundation for modelling even when underlying systems are fragmented.
- Generating realistic synthetic supply chain data to allow for prototype solution development and demonstration without access to proprietary data (although the latter is still required to properly validate that the solution meets specific customer needs and data patterns).

Part 7: Next steps

Our work in 2025 focused on research into the challenges facing supply chains and the potential for Safeguarded AI solutions.

In 2026, we start to make these solutions a reality – working with industrial partners and the experts in the Safeguarded AI programme to scope, plan and deliver pilots which pave the way for a large-scale roll-out of impactful and trusted AI solutions.

Specifically, we will:

1. validate the technical feasibility, operational safety, and impact of the most promising applications via a series of pilots with multinational firms.
2. stress-test assumptions, identify failure modes, and refine safety constraints in real-world industrial settings.
3. continue the systematic gathering, curation, and generation of representative supply chain data that accurately reflects the complexity and variability of the domain, to underpin modelling, experimentation, and solution development.
4. work with mathematical experts in the Safeguarded AI programme to further refine the approach to modelling and verifying the outcome of decisions in supply chains, including capturing key operational dynamics and constraints.

Through this we will build, test, and scale safeguarded AI solutions that are technically robust, practically feasible, and trusted and valued by industrial partners.

If you would like to be involved in our research or pilot projects, or learn more about our work, please contact:

- Dei Vilkinsons <d@hash.ai>
- Ciaran Morinan <c@hash.ai>