

Whitepaper

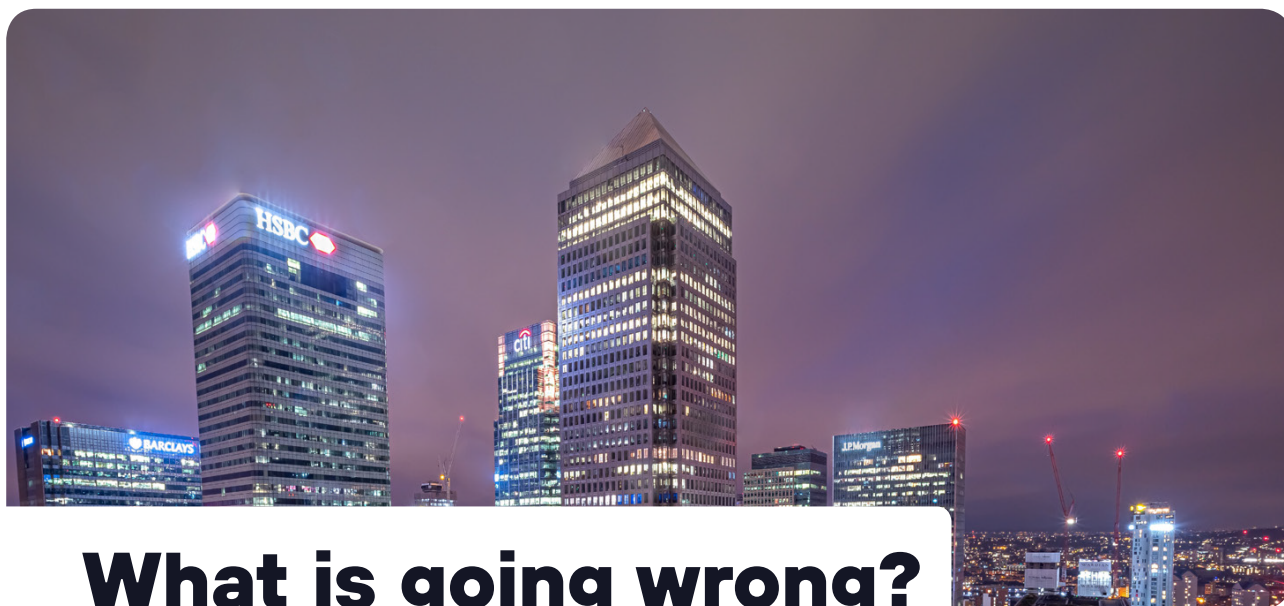
How Machine Learning is Enabling New Cost Levers in Post-Trade Operations



Contents

In this whitepaper, we discuss the challenges banks are facing when transforming their legacy operations — and how recent developments in machine learning are providing new ways to accelerate large-scale change and deliver the 'Zero-Ops' vision.

What is going wrong?	3
A new approach?	4
Google can do this right?	5
Solution	7
Case Study	9
Fixed income settlement	12
Prototype & Platform Design	14
Embracing Zero Ops	17



What is going wrong?

A decade on from the global financial crisis, many banks (particularly those in Europe) are still struggling with profitability and high cost-income ratios.

Alongside efforts to increase revenues, banks have made operational efficiency the main focus of their cost strategy— it is now widely accepted that the back office is key to unlocking the next level of cost savings.

Most banks have continued to pull traditional cost levers such as outsourcing, process excellence, and workflow management.

More recently, there has been a lot of focus on robotic process automation (RPA).

In many cases, the results have been underwhelming, with multi-year change programs failing to deliver the expected benefits. Operating costs remain stubbornly high, and in some cases, cost-income ratios are increasing.

A new approach?

As banks seek new efficiency levers they have started to explore how new technology can accelerate opportunity discovery and transformation efforts.

Much has been written about the potential benefits of new technology in financial services — strategies such as Ops 4.0 and Digital Transformation promise everything from significant productivity improvements to the emergence of new business models.

Not to mention industry-wide market infrastructure renewal enabled by Blockchain technology.

These strategies make a lot of sense on paper, but so far there are very few real-world examples of them being implemented successfully — and little evidence they can deliver anything more than superficial change.

One approach for opportunity discovery that is starting to gain traction is to use machine learning to mine the large quantity of communication data in an organization.

"I'm totally convinced that the battleground of banking is not the front office. The battleground is the back end."

Sergio Ermotti, CEO UBS



Google can do this right?

Products such as Google Duplex are starting to showcase the capabilities of deep learning for understanding and automating human conversation.

For conversations that are relatively bounded (like booking an appointment), Google can create very effective solutions as it has access to vast quantities of training data.

Within the enterprise, however, conversations are typically unbounded with only small quantities of data available for training.

For example, traders typically use highly specific language that can vary by product and from desk to desk.

Unbounded conversations require a different set of machine learning techniques that can deal with small data samples.

"90% of enterprise data is unstructured and a significant portion of this data is text in the form of emails, customer support chats and notes."

International Data Corporation

Solution

Re:infer combines unsupervised and supervised machine learning techniques to automatically detect concepts and intents in communication data, avoiding the need for complex heuristics.

Business users simply create a taxonomy for their data using a multi-stage learning process.

A taxonomy consists of business-relevant labels that are used to tag each conversation (together with confidence levels) in order to drive downstream processing.

In the case of post-trade operations, a taxonomy will typically describe a hierarchical processes model that identifies the function and exception type, e.g.

"Settlement > Fails > Counterparty Lack".

Taxonomies can represent any concept relevant to the business, for example:

- Event type (e.g. failure demand, trade lifecycle event, client inquiry...)
- Exception type
- Risk (e.g. GDPR violation, password sharing, elective corporate action instruction ...)
- Process /organizational unit
- Urgency
- Sentiment
- Entities (e.g. dates, currencies, instruments, prices, counterparties ...)
- Unique identifiers (e.g. ISIN codes, trade Ids, LEIs ...)
- ... etc

A key differentiating feature is that Re:infer only requires 15–20 data samples to detect new concepts with high confidence.

This is a significant step forward in the field of natural language processing (NLP) that offers two major business benefits:

- 1. Detection of high risk, low-frequency events.** Solutions that require 100s or 1,000s of data samples simply cannot detect these events.
- 2. Very short training cycles.** Other solutions require a lot of data and typically take weeks or months of effort to train a model. Training cycles can now be measured in days or hours, even for large data sets.

Once trained, a taxonomy can be deployed in two modes:

Offline — for analysis and opportunity discovery.

Online — for applying labels to conversations in real-time in order to drive downstream automation.

The team at Re:infer are at the forefront of research in this space and are actively working with a select group of banks who are early adopters of this technology.



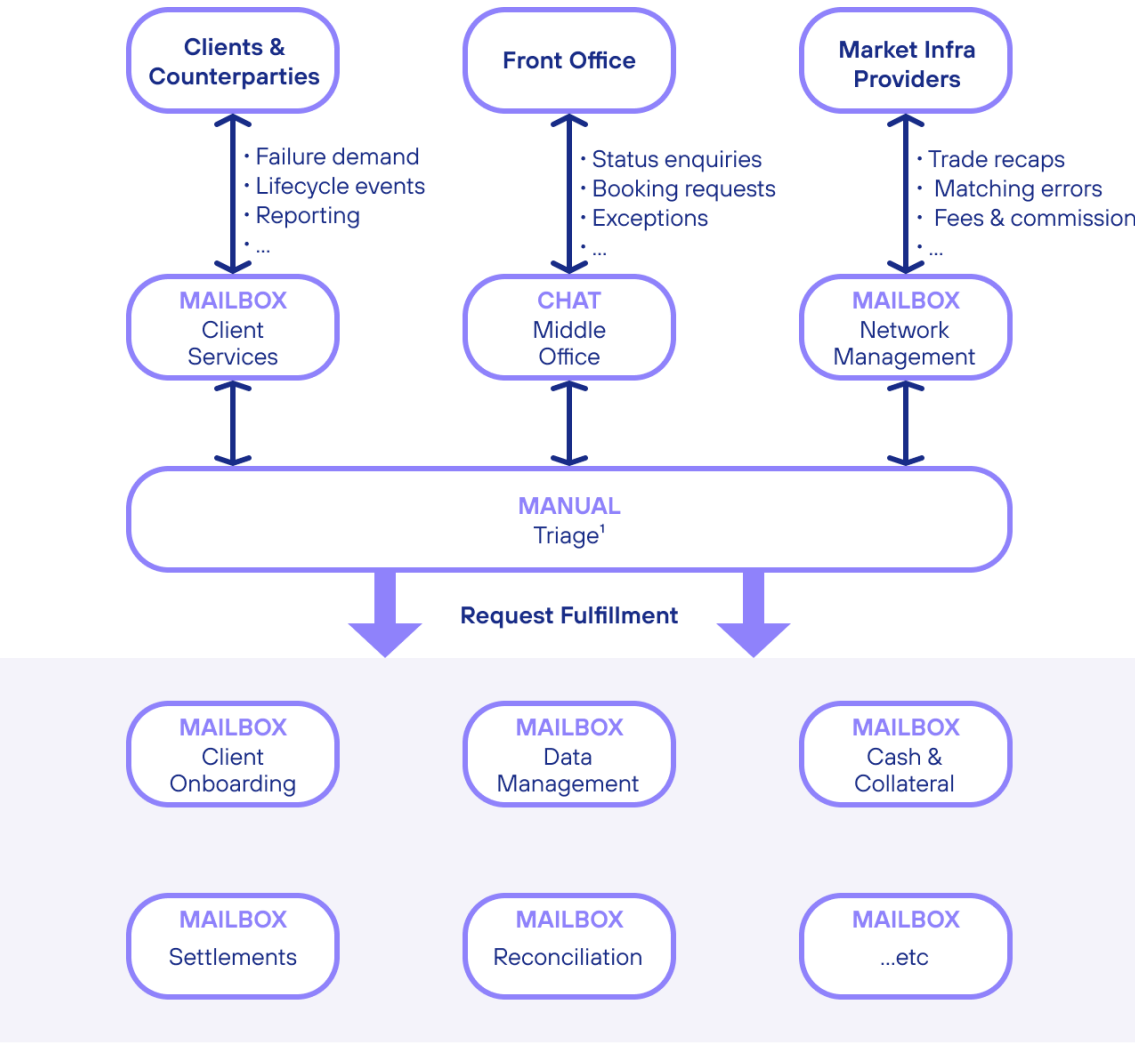
Case Study

In post-trade operations, Reinfer is being used to discover manual processes, quantify failures, highlight root-cause errors, detect incorrect reference data and provide a bridge to downstream automation.

The first step is to get a holistic view of the communication flows across an organization and understand how email is being processed.

This can be done quickly by building a broad taxonomy that describes the post-trade processes and course-grained exception types and applying it to the client's email data.

The following diagram shows the actual setup at a large European bank — all communication flows are handled manually and are entirely independent of the core IT systems and structured data flow.



¹Categorisation, prioritisation & routing

Source: reinfer.io

In this instance, the taxonomy was applied to 300 shared mailboxes — around 40% of the total email corpus for operations and over 5TBs of data.

Key observations:

A very large volume of historical data with a total of 3 billion emails and nearly 150 TBs of data stored on internal email servers, around 10% of which is operations data spread across 3–4,000 shared mailboxes.

A high volume of emails with around 500k – 1 million emails per day being handled across the middle and back office teams.

Manual triage with heavy use of shared mailboxes for manually categorizing, prioritizing and routing emails, accounting for around 5–10% of the effort in each operations team.

Communication is mostly internal with around 70% of emails being sent between internal teams and evidence of multiple handoffs between shared mailboxes.

Teams organized around mailbox structures rather than being aligned to a consistent model such as client/product/function.

Unbounded client communication channel with over 500 unique email addresses used by operations to communicate with a single client over a four week period.

Data duplication due to heavy use of distribution lists, causing data to be copied to multiple teams and individuals. In one case an external client cc'd over 20 separate distribution lists when sending an email.

Data deletion as 50% of operations email data is from the past 2 years, indicating a large volume of data is deleted after that period.

Inappropriate use of mailboxes for document management with some shared mailboxes having in excess of 20,000 subfolders.

Very poor search performance impacting the usability of workstations and degrading end-user productivity.

It is clear to see how deeply ingrained email and chat applications can become in large organizations. In this specific case, the operations team were using Microsoft Outlook as their primary tool for workflow, document management and enterprise search — none of which it was designed for.

Fixed income settlement

One area that was quickly identified as problematic in this case study was EMEA fixed income settlements and the related treasury functions.

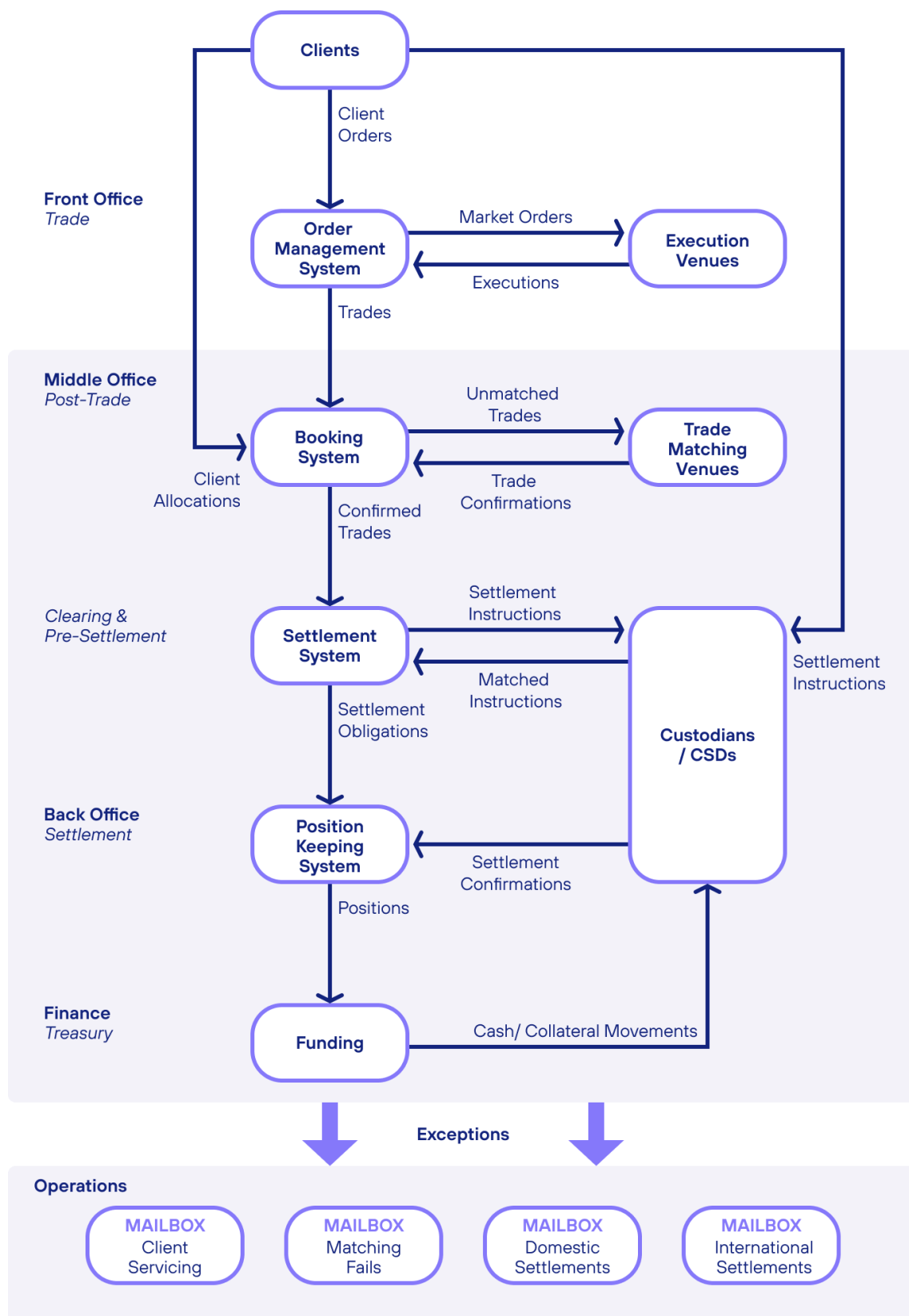
The fixed income operation supports bond and repo settlement for multiple central securities depositories (CSDs):

- U.K. Gilts and Money Markets (Debt Management Office)
- Domestic Settlement (Domestic CSDs via SWIFT)
- International Settlement (Clearstream and Euroclear ICSDs)

The functional scope managed by the operations team covers all post-execution functions including trade matching, allocation, confirmation, settlement instruction, settlement confirmation, and position keeping.

Additionally, cash and stock positions are fed to the treasury team to facilitate funding of the firm's activities (e.g. ensuring the bank has sufficient cash and collateral to cover its obligations, managing risk-weighted assets etc).

All queries and exceptions across these functions are handled by the operations team using shared mailboxes.



The diagram above shows the front-to-back process being managed.
Front-to-back fixed income process at a large European investment bank.



Prototype & Platform Design

The operations team consisted of around 70 full-time roles (excluding treasury operations), around 50 of which were located offshore. The annual FTE costs for a team of this size is around USD 3–4 million.

Client queries are raised to the operations team through client servicing mailboxes, where they go through a manual triage process before being routed to processing teams.

Process exceptions are raised to operations through functionally aligned mailboxes e.g. matching fails. In some cases, these exceptions are delivered to the operations teams automatically, but for the most part, operational staff have procedures to pull exception reports from the underlying systems at pre-defined intervals.

Internal queries are generated by multiple teams including the front office, business / COO, treasury, reconciliations etc. These are handled by the operations teams using additional shared mailboxes.

Average trade volume was around 10–15,000 per day, or 2.5–3.5 million per year. These volumes are very small in comparison to Equities and FX trade volumes.

Analysis of the fixed income mailboxes helped identify (and quantify) specific issues that needed to be addressed: Client behaviors that cause operational issues e.g. lack of cash/stock to deliver, late settlement instructions etc. These issues were mostly rectified by the business working together with the clients, but in some cases led to clients being off-boarded.

A significant number of inter-company trades failing settlement due to inventory management issues, specifically a lack of cash/stock to deliver on the intended settlement date (ISD). These issues led to a new initiative to improve the real-time

accuracy of main-firm positions and front-to-back visibility of positions through shared infrastructure.

Sub-optimal collateral and funding processes due to inaccurate inventory and matching/settlement fails not being included in short term funding ladders. These issues led to a large volume of status inquiries over chat and email from the treasury trading desks, finance, and collateral teams.

Ineffective trade matching process on repo flows with most errors only being identified through the settlement instruction process – much later in the trade lifecycle than is necessary. For the most part, trade matching issues are managed manually over email with the counter-parties and custodians.

On average around 8-10% of trades settle late i.e. settlement occurs later than the intended settlement date (ISD) – despite most settlement instructions being matched before on time.

Late settling trades cause operational backlogs contrary to the widely held belief that operational processes are volume insensitive. Ordinarily, around one-third of late settling trades are settled within 1–3 days, 4–10 days and >10 days respectively, but during times of stress the average time to remediate late settling trades increases.

Custodians send over 3,000 error messages per day or around 500k messages over a 6 month period. Around one-third of these were booking issues, one third relating to settlement and inventory management problems, and one third largely due to processing latency or other types of unnecessary waste.

Trade lifecycle events sent over email present risk, for example, trade confirmations/ allocations, corporate action instructions etc. These would typically go through a manual triage process and then lead to manual data entry into core IT systems or market matching utilities. This led to an improvement program to either move lifecycle events onto systemic flows (e.g. FIX) or to automate the triage process and call downstream robots for processing.

Sub-optimal process design e.g. when a client lacks the cash/stock to deliver the original process was to let it fail settlement. The process was subsequently changed so that the trading desk asked the client for partial settlement and covered the shortfall through the repo desk, thereby avoiding settlement failure and improving the accuracy of positions.

Poor organizational design creating a high number of handoffs between operational teams (e.g. unmatched settlement instructions sent to the settlement fails team) would

always create a request to the confirmations team to obtain the trade details agreed with the client, and then if any trade amendments are required that creates further requests to the middle office and trade support teams. These handoffs were eliminated by re-organizing the teams and mailbox structures around front-to-back product flow, so each team member could manage a settlement fail independently.

The above issues were all identified from communication data using Reinfer, and led to a broad program of targeted improvements that encompassed:

- Review of problematic clients
- Bug fixes to underlying systems
- Shared front-to-back views of trade status and positions ("Trade Tracker")
- Process improvements and redesign
- Automation of manual triage process
- Automatic triggering of downstream robots
- Automatic case creation in downstream workflow systems
- Phasing out of manual reporting
- Development of intelligent alerting and risk indicators

As the initial analysis approach was entirely data-driven, the impact of these changes could be measured against historical levels.

It is realistic to deliver 30% efficiency savings within a reasonably short timeframe, purely by focusing on communication flows. Further savings are possible by focusing on other sources of unstructured data and combining transactional data — the latter being a necessary step to develop more advanced solutions such as anomaly-based process management.

Embracing Zero Ops

Early indications are that machine learning can help unlock the complexity of the back office environment — thanks to recent advances in deep learning, it is now possible to mine unstructured data at scale and quickly identify the inefficiencies and manual processes that have built up over time.

Interestingly, these advancements start to shape a future vision for operations that is radically different to today. Operations is fundamentally a risk function. It manages risk from the point of trade execution through to settlement. Any risks that occur post-execution are owned and managed by operations until they are resolved.

Granted there are a lot of responsibilities — bookkeeping, regulatory reporting, data management, reconciliations etc. — but all of these functions are fundamentally about managing risk.

In practice, however, operations have largely become a processing function. A lot of management attention is typically focused on operating models, process execution and reducing the unit cost of labor.

The combination of manual processing in a complex environment is problematic. Risk events are not infrequent and the vast majority of them are driven by manual errors, poor process design or inadequate procedures.

In principle, there is no reason why processing and exception management cannot be fully automated.

This must be the core focus of any technology-enabled strategy — to transform operations from a process function to a risk function by eliminating manual processing and exceptions.

In the future, operations will have much higher levels of automation than exists today, and a significantly smaller footprint. Anomaly detection will allow operations to 'shift-left', focusing on prevention rather than reacting to exceptions after they occur.

This is the vision of Zero-Ops. Highly automated. Low cost. Proactive. Risk-focused.

Machine learning will play a significant role in realizing this vision.

Bring artificial intelligence into every conversation

Discover, measure, and automate the processes hidden in your communications data to drive efficiency and revenue.
See Re:infer in action today.

Re:infer

Authors

Martin Swanson
Enterprise Architect at UBS
LinkedIn: [linkedin.com/in/martinswanson/](https://www.linkedin.com/in/martinswanson/)

Contributors

Dr Ed Challis
Co-Founder and CEO, Re:infer
LinkedIn: [linkedin.com/in/edwardchallis/](https://www.linkedin.com/in/edwardchallis/)

Stephen Mackintosh
CCO, Re:infer
LinkedIn: [linkedin.com/in/stephenmackintosh/](https://www.linkedin.com/in/stephenmackintosh/)

Online

Email : info@reinfer.io
Website : www.reinfer.io
Contact : www.reinfer.io/contact

Social

LinkedIn : [linkedin.com/company/re-infer](https://www.linkedin.com/company/re-infer)
Twitter : twitter.com/reinferio