# Exploring the applications of Artificial Intelligence in Payment Market Infrastructures

PAYMENTS CANADA

CGI

# Exploring the applications of Artificial Intelligence in Payment Market Infrastructures

May 2019

**CGI**

250 Yonge Street, Suite 2000,
Toronto, ON  M5B 2L7

416-363-7827
cgi.com/canada

**Payments Canada**

Constitution Square, Tower II
350 Albert, Suite 800
Ottawa, ON  K1R 1A4

22 Adelaide St W, 24th Floor
Toronto, ON M5H 4E3

613-238-4173
payments.ca

## Authors

**CGI**

**Ainsley Ward**
Vice President, Consulting Expert
Strategic Payments Architect

**Neel Mehta**, FRM
Director, Consulting
Liquidity Expert

**Payments Canada**

**Neville Arjani**
Director of Research, Payments Canada

**Segun Bewaji**
Senior Economist, Payments Canada

# Smarter payments starting at the centre

## While the application of artificial intelligence (AI) is embedded within daily enterprise operations across a range of industries—for example, entertainment, healthcare, and media communications—the application of AI within the payments industry is still in its infancy outside of the FinTech segment.[1]

Common use cases include anomaly detection algorithms to support fraud identification, anti-money laundering or sanctions screening, and enabling expedient and accurate payment authorization. Another use case includes streamlining and automating manual tasks in the accounts receivable and accounts payable functions. (e.g. using algorithms that leverage existing customer data to match incoming electronic funds transfers with invoices and remittance information sent separately by the customer via fax or email). Some businesses are also turning to natural language processing to enable virtual assistants to address common customer inquiries, including payment initiation requests, thereby allowing service representatives to spend more time on complex client matters.

Importantly, in payments as in other industries, AI adoption is 'human-centric'. While not intended to replace humans outright (which is a common concern relating to AI[2]), in this case the technology is left to do what it does best: to identify complex patterns in data by navigating reams of records in a fraction of the time it would take any human, and applying this learning to new records for the purpose of prediction or classification. With this, employees are able to leverage their comparative advantage, including assuming responsibility for more complex tasks where inputs and outputs are more variable in nature, as well as applying judgment in validating the output of the AI application, and even providing deep domain knowledge to create the AI application in the first place.

[1] Use of the term AI is intended to be broad in this report, encompassing applications of machine learning, deep learning, and natural language processing, for example.

[2] See, for example, Fortune. 2019. "A.I. expert says automation could replace 40% of jobs in 15 years", accessed on March 22, 2019 at http://fortune.com/2019/01/10/automation-replace-jobs/.

With this said, there appears to be a gap in that, while AI implementations are beginning to take shape at the customer-facing end of the payments value chain, there is limited discussion of implementation at the supporting back-end of the payments value chain—defined in this report as the Payment Market Infrastructures that constitute the wholesale payments system—which brings together financial intermediaries and financial market infrastructures such as Payments Canada, participating financial intermediaries, and technology service providers like CGI.

The remainder of this report focuses on the merits of AI implementation within the national wholesale payments system, and contemplates potential use case applications with the hope of generating industry reflection and subsequent dialogue.

## The wholesale payments system as a critical component of the financial system and broader economy

When operating well, the national wholesale payments system goes largely unnoticed by the general public. However, it is a critical component of the financial system and underpins virtually all economic and financial activity in the country. Failures in this system can have a dramatic negative effect on economic activity, while at the same time optimization and openness can improve economic growth. This infrastructure encompasses a vast web of centralized and decentralized hardware and software, network and communications technology, information security protocols, and a host of rules, standards and procedures governing activity among connected financial intermediaries (system participants). Participants leverage the system to transfer monetary value between transaction accounts, where these accounts may be proprietary to the participant, or instead maintained on behalf of clients. Monetary transfers between participants are derived from daily economic and financial activity. House purchases, business-to-business commercial activity, interbank transfers, mergers and acquisitions, cross-border commercial transactions, foreign exchange trading, or really any payment of considerable size or importance and which has an element of time-sensitivity and a need for immediate irrevocability are examples of transactions supported by the wholesale payments system. Not to mention that the wholesale payments system also serves as a platform for the daily implementation of monetary policy in most countries, including Canada. Put plainly, the wholesale payments system is the beating heart of any economy.

It therefore follows that the daily value transferred via the wholesale payments system is significant. In Canada, for example, the Large-Value Transfer System (LVTS), which is owned and operated by Payments Canada, clears roughly $200 billion in value each day. This translates to clearing the equivalent of Canadian Gross Domestic Product (GDP) every nine business days. Given the size and status of payments clearing through the wholesale payments system, and the system's importance to financial system soundness and efficiency, wholesale payments systems are subject to considerable oversight and regulation. The scope of oversight is broad, covering financial risk (e.g., credit and liquidity risk), operational risk (e.g. cyber and security risk), and overall business risk and resiliency. Indeed, in many countries, ownership and operation of the wholesale payments system is the domain of the central bank. In Canada, while the LVTS is owned and operated by Payments Canada, participating financial institutions' LVTS settlement accounts are held with the Bank of Canada, which is also the designated oversight authority of the LVTS.[3]

> **"In Canada, for example, the Large-Value Transfer System (LVTS), which is owned and operated by Payments Canada, clears roughly $200 billion in value each day. This translates to clearing the equivalent of Canadian Gross Domestic Product (GDP) every nine business days."**

[3] For more information on the LVTS, see N. Arjani and D. McVanel. 2006. *A Primer on Canada's Large Value Transfer System*. The paper is available from https://www.bankofcanada.ca/wp-content/uploads/2010/05/lvts_neville.pdf.

## The wholesale payments system is a prime candidate for AI implementation

The wholesale payments system remains largely untapped in contemplating the merits of AI application. Ward and Mehta (2018) and Triepels et al (2017) are recent examples of research in this area; otherwise the literature is fairly scant.[4]

In our view, the wholesale payments system represents a strong use case for AI-powered applications due to its fundamental importance in relation to:

1. national economic well-being;
2. material operational costs; for instance, requiring ongoing coordination of human and technology resources across the treasury, cash management and middle office functions of participant financial intermediaries;
3. the high degree of manual and repetitive effort in some pockets of the wholesale and commercial payments business — as the global payments industry faces a looming talent shortage, the opportunity to relieve staff of repetitive and manual tasks by leveraging AI in favor of more complex responsibilities is a very strong proposition;
4. large volumes of well-structured and high-frequency data are generally available from both centralized and decentralized systems comprising the wholesale payments system, and can date back far enough to capture behavior across different phases of the economic and financial cycle.
5. both market protocols and preferred behaviours among system participants, strong patterns seem to emerge in the wholesale payments data at a variety of levels and frequencies, including quarterly, monthly, daily and even on an intraday basis (see Figures 1 & 2 below).

The following section takes this discussion a step further by contemplating, at a high level, potential use cases for AI implementation in the wholesale payments environment. While not intended to be exhaustive in terms of scope and content, the intent is for these cases to engender thoughtful reflection and dialogue within the industry and prompt further exploration.

Moreover, while subsequent discussion focuses predominantly on the potential application and utility of AI in the wholesale payments environment, it is, of course, critical that implementation of this technology is carried out in compliance with regulatory direction around system safety, security and information privacy to protect the interests of users and broader stakeholders. As mentioned, the smooth daily functioning of these systems is critical to national economic well-being. This would perhaps preclude use of so-called "black box" algorithms in favour of algorithm transparency and explainability. Such considerations are beyond the scope of the current paper, and are recommended as areas of further research.

[4] Ainsley Ward and Neel Mehta. 2018. Balancing Liquidity and Risk in Modern Payment Systems: Use of AI-controlled dynamic periodic net settlement mechanisms in real-time payment market infrastructures. CGI White Paper. Available from: https://www.cgi.com/sites/default/files/balancing-liquidity-risk-modern-payment-systems.pdf.

Ron Triepels, Hennie Daniels, and Ronald Heijmans. 2017. Anomaly Detection in Real-Time Gross Settlement Systems. Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS 2017) – Volume 1, pages 433 – 441. Available from: http://www.scitepress.org/Papers/2017/63330/63330.pdf.

**Figure 1 – Demonstration of regular patterns in daily LVTS clearing volume**

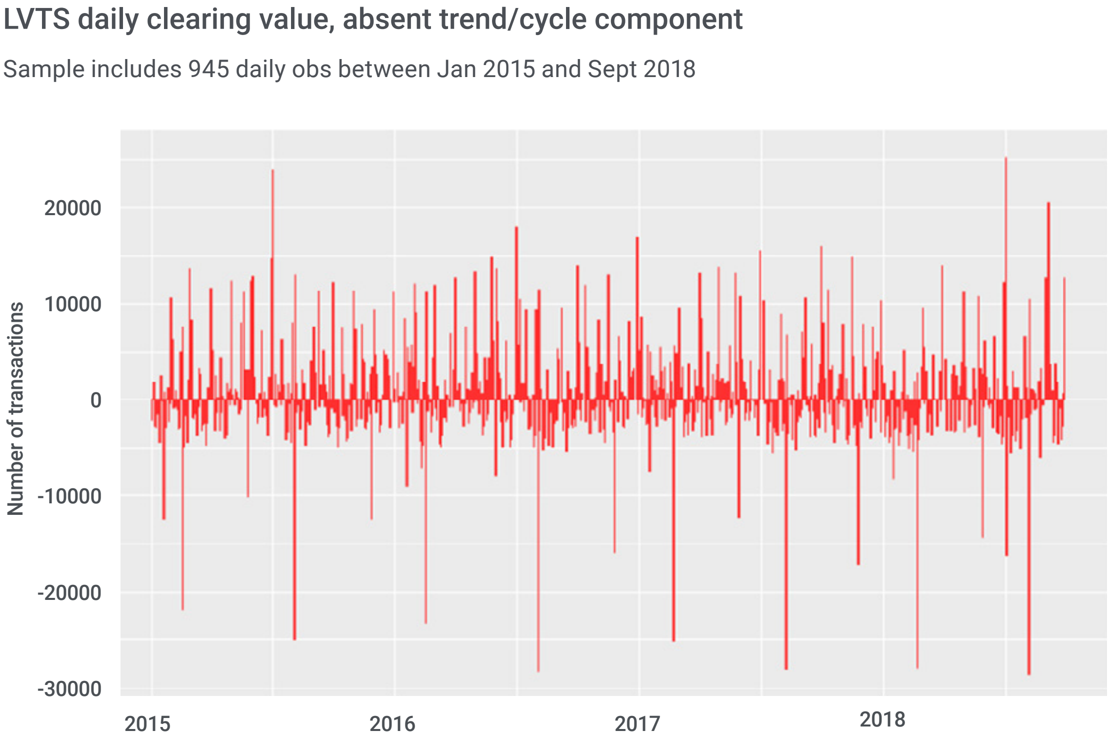


**Figure 1** demonstrates LVTS daily clearing volume with trend and cycle effects removed, for the period between January 2015 and September 2018. The Civic Holiday in Canada, as well as U.S. National Holidays figure prominently, year after year. Moreover, elevated volume at month-end and month-beginning are a regular feature in the data, as are regular spikes on the third-Wednesday of each quarter-end, likely relating to particular financial market activity around these dates, e.g., settlement of 3-month banker's acceptances.

**Figure 2 – Demonstration of regular patterns in intraday LVTS clearing volume**

### Intraday profile of LVTS clearing volume: 10 minute intervals

Based on 1701 days from 2012M1 to 2018M9; Shaded area reflects distance between 10th and 90th percentiles

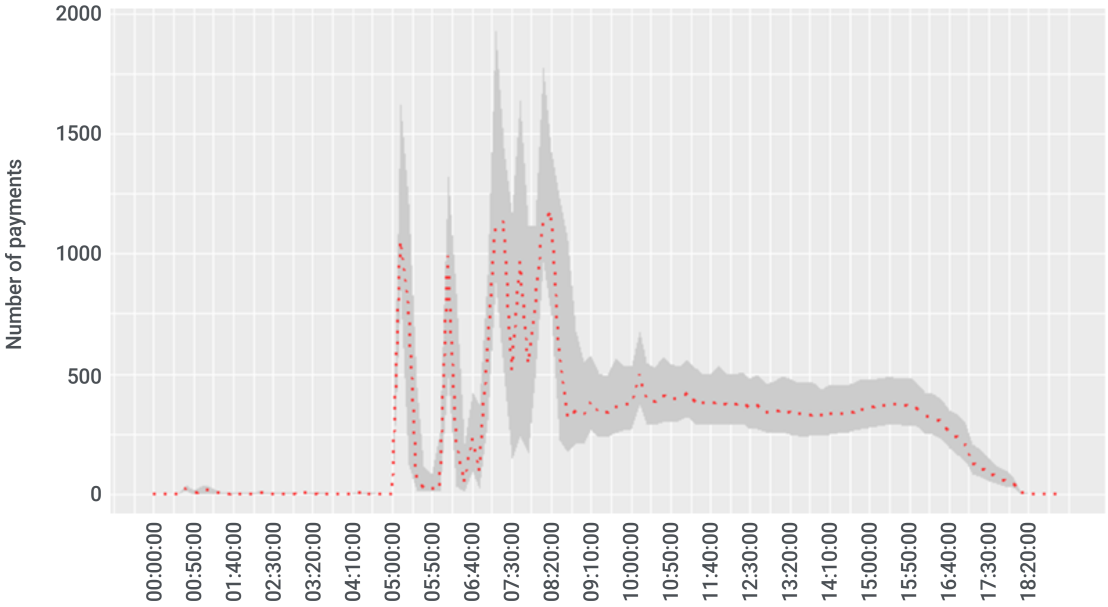


**Figure 2** demonstrates typical LVTS clearing volume at 10-minute intraday intervals over a near 7-year period between January 2012 and September 2018 (1701 operating days in total). What stands out from this graph is that, for the vast part of the 18-hour LVTS operating day, one can expect clearing volume in any 10-minute interval to fall within a range of a few hundred payments. This also means that deviations from typical payment submission behavior, such as an operational disruption that causes a system participant to fall behind its regular submission pattern, can generally be identified quickly.

# Exploring potential use cases of AI implementation within the wholesale payments environment

This thought paper marks the beginning of an exploration into the merits of AI-powered applications in the wholesale payments system. We identify four use cases where AI and predictive analytics could be applied to solve market problems or reduce repetitive manual work while improving overall efficiency.

- **Liquidity savings and optimization**—applying functions to control central processing in line with overall market goals, increasing efficiency and lowering the liquidity burden of participants
- **Cash flow forecasting**—allowing the central market to interact with participant systems and propose effective timing for payments to maximise synchronization
- **Gridlock avoidance**—providing the market participants with advice and direction on appropriate liquidity levels to maximise market efficiency
- **Surge pricing policies**—implementing pricing policies that drive positive behaviour by the participants

Readers requiring further background on the wholesale payments system prior to evaluating the following use cases are encouraged to review pages 8 to 10.

# A brief primer on wholesale payments system design

## The Real-Time Gross Settlement (RTGS) model underpins virtually all wholesale payments systems in the world as a means of appropriately containing settlement risk among participating financial intermediaries.[5,6]

Central to wholesale payments system design is the notion of intraday liquidity, which refers to the ability of a system participant to meet its payment obligations in a timely manner as they become due—where timing is driven by proprietary, client and sometimes even regulatory needs. In a RTGS model, each transfer of value between participants involves the immediate and irrevocable transmission of balances from the sending participant's RTGS settlement account to the receiving participant's RTGS settlement account, where settlement accounts are maintained by the Central Bank. That is, a participant must maintain sufficient funds in its settlement account to meet its payment obligations at the time that they become due.

[5] The Bank for International Settlements offers further information on the motivation and evolution of RTGS systems around the world through the following papers.

Bank for International Settlements. 1997. *Real Time Gross Settlement Systems*. CPMI Papers. No. 22. Available from: https://www.bis.org/cpmi/publ/d22.htm.

Bank for International Settlements. 2005. *New Developments in Large-Value Payments Systems*. CPMI Papers. No. 67. Available from: https://www.bis.org/cpmi/publ/d67.htm.

[6] For more information on the costs and risks in wholesale payments systems, see the following papers.

William R. Emmons. 1997. *Recent Developments in Wholesale Payments Systems*. Federal Reserve Bank of St. Louis Review. November. Pages 23 – 43. Available from: https://files.stlouisfed.org/files/htdocs/publications/review/97/11/9711we.pdf.

Allan Berger, Diana Hancock and Jeffrey C. Marquardt. 1996. *A Framework for Analyzing Efficiency, Risks, Costs, and Innovations in the Payments System*. Journal of Money, Credit and Banking. Vol. 28, No. 4, Part 2: Payment Systems Research and Public Policy Risk, Efficiency, and Innovation. November. Pages 696 – 732.

There are typically three main sources of intraday liquidity in the RTGS environment, including:

- the stock of own settlement balances (or reserves) the participant holds on account at the central bank (the settlement agent);
- incoming payments received from other participants, which increase the participant's settlement balance; and,
- intraday credit provided by the central bank, which is typically secured by eligible collateral securities.

Each of these sources contributes to the capacity of a participant to transmit value over the network expeditiously. Each source, however, also entails a different 'cost' for participants. For instance, incoming payments are, for all intents and purposes, a costless source of intraday liquidity, whereas central bank credit is comparatively expensive and depends on the cost of collateral which is time-varying and state dependent (e.g., normal market conditions vs market stress conditions).[7] The cost of collateral facing a participant is also influenced by other, participant-specific features, such as business model and depth of market operations.

Naturally, participants prefer to rely on the less expensive source of liquidity to fund payments activity, other things equal. As a result, participants would tend to rely on incoming payments to fund their outgoing payments, depending on the cost of delaying customer payments. That is, different payments have varying time-sensitivity: some payments can be delayed for hours while others must be processed immediately upon receipt of instruction, depending on the preferences and needs of customers. As a result, participant behaviour will reflect not only the cost of liquidity, but also the costs associated with payment delay, subject to the (structural) constraints imposed by the wholesale payments system.

Put differently, in a basic RTGS set-up, all payments must be fully funded on a gross basis (i.e., no netting), and so such a system is liquidity-intensive and potentially costly. Therefore incentives are established for participants to conserve liquidity. As noted, incoming payments are a relatively inexpensive source of liquidity, and so participants predictably would withhold (or delay) non-time-sensitive payments while awaiting incoming funds. This is likely to lead to adverse outcomes, including, at the extreme, system gridlock, with severe negative externalities for the system and for the wider economy.[8]

Minimizing the collective cost of liquidity and payments delay, while continuing to appropriately contain operational risk, is fundamental to wholesale payments system design. Liquidity Savings Mechanisms (LSMs) are a means of improving outcomes for this problem. LSMs aim to economize on costly liquidity in an RTGS system, by encouraging timely submission of payments to manage

[7] Intraday credit not paid back by the end of business day is sometimes rolled over into an overnight credit arrangement, which would entail an interest charge (subject to monetary policy parameters) in addition to being secured by eligible collateral.

[8] For a theoretical treatment of the intraday liquidity management challenge and how delay and liquidity costs might underpin participant behavior, see Morten Bech and Rod Garratt. 2003. *The Intraday Liquidity Management Game*. Journal of Economic Theory. Vol 9, Issue 2. April. Pages 198 – 219.

For an empirical treatment for the US and UK systems, respectively, see the following papers.

James McAndrews and Samira Rajan. 2000. *The Timing and Funding of Fedwire Funds Transfers*. Economic Policy Review, Vol. 6, No. 2. July. Available at SSRN: https://ssrn.com/abstract=888772

Christopher Becher, Marco Galbiati, and Merxe Tudela. *The Timing and Funding of CHAPS Sterling Payments*. Economic Policy Review, Vol. 14, No. 2. September. Available at SSRN: https://ssrn.com/abstract=1141340 payments delay to an acceptable level. The term “LSM” is general in nature, and refers to any mechanism that supports better coordination of incoming and outgoing payment flows between Lynx Participants, thus reducing the likelihood in big intraday dislocations and heavy reliance on costly intraday credit. Examples of LSMs include central queuing with payment offsetting, intraday throughput requirements, net sender limits, and time-varying transaction tariffs. Relative to a basic RTGS set-up, such modifications would improve liquidity recycling (reduce liquidity hoarding), decrease the need for intraday credit extension from the central bank, and reduce payment delay, and so provide for satisfactory collective and individual outcomes. BIS (2005) provides an excellent overview of LSMs and their application to RTGS systems in aiming to incentivize appropriate behaviour among system participants.

AI Case Study #1:

# Liquidity Savings and Optimization

A common example of LSM as mentioned earlier is the use of a central queue that performs netting of queued payments according to some deterministic algorithm. By netting queued payments, the LSM helps to economize on the aggregate liquidity needed to clear the payments, which also lends to reduced payments delay and enhanced payments processing efficiency.

The “complex” algorithms employed by RTGS systems today may be parametrized in accordance with a number of factors, including frequency of netting, type of netting (e.g., bilateral and/or multilateral netting), and respecting (or bypassing) ordering and prioritization of payments. Partial netting is even quite common, where the LSM might identify a subset of payments that satisfy available liquidity supply, even though the release of all queued payments may not. Notwithstanding, once these parameters are established, the algorithm will then continue to run deterministically, day in and day out, and under alternative market scenarios, performing the same function at determined frequency (e.g., every 10 or 15 minutes). Any adjustments to these parameters, and therefore the functioning of the algorithm itself, remains a manual process under the authorization of the system operator or central bank. This means that the operator must constantly monitor payments system performance and external market conditions and make adjustments to parameters as needed with the hope of achieving optimized liquidity usage for the system.

Ward and Mehta (2018) challenge this logic, and argue that improvements to payments processing efficiency could be achieved by an AI-powered LSM that relies on system and market information and adapts the frequency of the central queue netting algorithm based on observed conditions. For instance, where payment traffic is fast flowing, and incoming and outgoing payments of system participants are well matched to the extent that few payments enter the queue, the algorithm would recognize such conditions and adjust automatically to run less frequently and thereby enhance netting power.[9]

In this context, one wonders whether AI can enable RTGS central queuing algorithms to shift further focus from “liquidity saving” to “liquidity optimization” while drastically limiting manual intervention needed by the system operator.

[9] This claim is based on an assumption that netting efficiency will be relatively greater the more payments that are queued centrally when the matching algorithm runs.

## How could this work?

Imagine the central queue as being operated by an 'autonomous agent', powered by AI and charged with maximizing an objective function defined by the system operator (perhaps in collaboration with participants and regulatory bodies), where this function maps the arguments of both liquidity and delay into some system-wide utility value. For example, low liquidity and low delay could translate into higher utility value, which is preferable. The function would serve as a guidance system for the agent, where the goal would be to operate the central queue by choosing when and how to process queued payments in a way that maximizes the value of this function over the long term. Preferences of the system operator over liquidity and delay would be embedded in the function itself, where additional constraints could also be imposed on the agent's problem through factors such as payments prioritization / timing restrictions, as well as other participant needs and preferences.

With the optimization problem in place, the autonomous central queuing agent is placed in its environment, ready to learn and adapt. Day after day, month after month, year after year, and under various system scenarios, the agent is faced with queued payments for processing, and moment after moment it must scan the system and the market environment and decide on a plan of action that contributes to the greatest utility level. Importantly, the agent must also factor in how a decision taken in one moment might impact prospective utility in subsequent moments as well.

Especially in the beginning, the agent is completely unaware of the environment it is in. By trial and error, and over reams of payments and market conditions data, it learns over time what works and what doesn't work in regard to utility maximization.[10] Over the long term, and to the extent that market conditions tend to follow cycles, the agent is able to acknowledge with greater (but not complete) certainty the system and market conditions at a given time, and becomes more efficient in its decision making just as a human might.

Critically, and unlike the current RTGS environment, the agent operates in a nondeterministic way and without need for regular intervention by the system operator. Put differently, just as humans learn to interact in uncertain environments, where they learn over time what adjustments to behaviour are needed to maximize well-being, so too would the central queue agent. Timely adaptation to changing market conditions is needed, with recognition that a "one size fits all" action strategy under all conditions is not optimizing over the long run. And with this we move from an objective of "liquidity saving" to one of "liquidity optimization."

[10] For the sake of safety and efficiency, nothing would seem to preclude much learning to take place on historical transaction records and prior to introducing the autonomous agent to a production environment.

AI Case Study #2:

# Cash Flow Forecasting

The intersection of payments and intraday liquidity management is a key area of focus in minimizing the trade-off between payment processing and liquidity efficiency.

AI's key differentiator is the speed in predicting future net liquidity balances at a single point in time in order to predict maximum net cumulative payment outflows proactively. Traditional methods of end of day cash flow forecasting have involved the intensive use of spreadsheet modelling to incorporate payment flow patterns in a static snapshot with parameter calibration occurring on an ad hoc basis. Centralizing expected inflows and outflows from disparate systems remains a key challenge, where the manual approach of aggregating datasets has high probability of human error. More importantly, behavioural pattern overlays such as seasonality or payment flow drift are often based on cash manager's intuition rather than statistical inference. The lack of rigor in accurate modelling of cash flows partly stems from the current environment of abundant liquidity and low interest rates, which do not provide sufficient cost pressure to persuade financial institutions to invest in dynamic and accurate cash flow positioning techniques in order to avoid excess collateral buffers.

## Key drivers for improving cash flow management

The global drive towards central clearing and real time settlement infrastructures coupled with macroeconomic headwinds and monetary policy tightening has increased the need to not only manage increased collateral requirements but also source scarcer and costlier sources of intraday liquidity.

The accuracy and robustness of forecasting tools is a critical enabler to banks optimizing their liquidity reserves efficiently in a rapidly changing liquidity landscape.

Augmented Intelligence can play a vital role in aiding human decision making for value add what-if scenario generation in forecasting activities while taking over operationally intensive payment throttling tasks.

## Payment Velocity

Instant payments schemes, 24x7 mobile payments and RTGS central market infrastructures are increasing liquidity velocity as clearing and settlement cycles become almost instantaneous. Although banks can benefit in offering innovative product suites to take advantage of diminishing cash conversion cycles, the timing and coordination of large value payments has large implications on overall costs of maintaining collateral buffers, stabilizing payment infrastructures and meeting robust regulatory requirements. As the velocity of payments increases, the requirement for real time liquidity management becomes more pressing.

Specifically, a comprehensive payment throttling strategy gains more importance in liquidity intensive RTGS payment infrastructures. In collateralized credit regimes, intraday liquidity is a function of pledged collateral (less the value of applicable haircuts) and the sum of payment inflows and outflows. The trade-off between payment timing and liquidity efficiency is partially mitigated through the use of liquidity savings mechanisms that make use of First-In First-Out (FIFO) bypass and gridlock busting algorithms[11] to offset payments. However, the problem of payment synchronization exists as banks can choose to be receipt reactive in order to conserve liquidity in a macro economically constrained environment.

AI's computational power and processing speed has important ramifications for accurate and rapid cash flow forecasting, where swift decision making can be a compelling differentiator in reducing receipt- reactive behaviour, and in minimizing the stock of liquid assets and associated collateral costs to downstream products, thereby enhancing competitive franchise pricing and positioning.

## Cost implications of tightening monetary policy

The current macroeconomic context and monetary policy implication offers a compelling rationale to conserve liquidity. Tepid US growth prospects amidst trade war concerns potentially has a knock on effect on dampening investment and consumption spending in Canada. Moreover, the increased probability of a hard Brexit raises the fallout of a costly windup or novation of $427tn derivative market, of which approximately 90 per cent is cleared in the United Kingdom. Finally, quantitative tightening through asset purchases and interest rate hikes is expected to drain liquidity from markets and increase competitive pressure in obtaining longer duration and stable sources of liquidity. Specifically, higher interest rates increase the explicit and implicit costs of maintaining collateral buffers for intraday liquidity usage. The explicit costs emanate from liquidity reservation where collateral is pledged in payment infrastructures and cannot be used for repo facilities, which have traditionally reduced funding costs by 30 – 60bps.[12]

The implicit cost is significantly larger, as banks forego utilizing liquidity in their preferred client portfolios that maximize return on economic capital. Our estimates of the interest rate differential between a liquid asset portfolio and preferred client portfolio can range between 100 – 145bps, potentially reaching even higher at 200bps in a rising rate scenario where treasurers invest in shorter duration securities to minimize MtM loss in a steepening yield curve.[13] As cost per unit of collateral increases

[11] See Use Case 3 for more detail on gridlock busting

[12] CGI internal point of view.

[13] CGI Internal point of view.

and collateral requirements increase in cover all RTGS systems, costs become too punitive to ignore. Therefore, cash flow forecasting becomes essential to minimizing buffers for the smooth and cost effective processing of payment traffic.

## Dimensions of cash flow forecasting

Cash flow forecasting typically straddles across multiple business objectives and dimensions in banks. Short term liquidity planning incorporates forecast visibility on a daily basis for a period up to 1 month while Liquidity risk management utilizes cash forecasts as early warning indicators for future liquidity stress. Other dimensions include forecasting in the medium and long term for interest and for debt reduction and quarter end snapshot forecasts for covenant and key date visibility. This use case focuses on leveraging AI in short term liquidity planning with particular emphasis on intraday timing and coordination of cash flows.

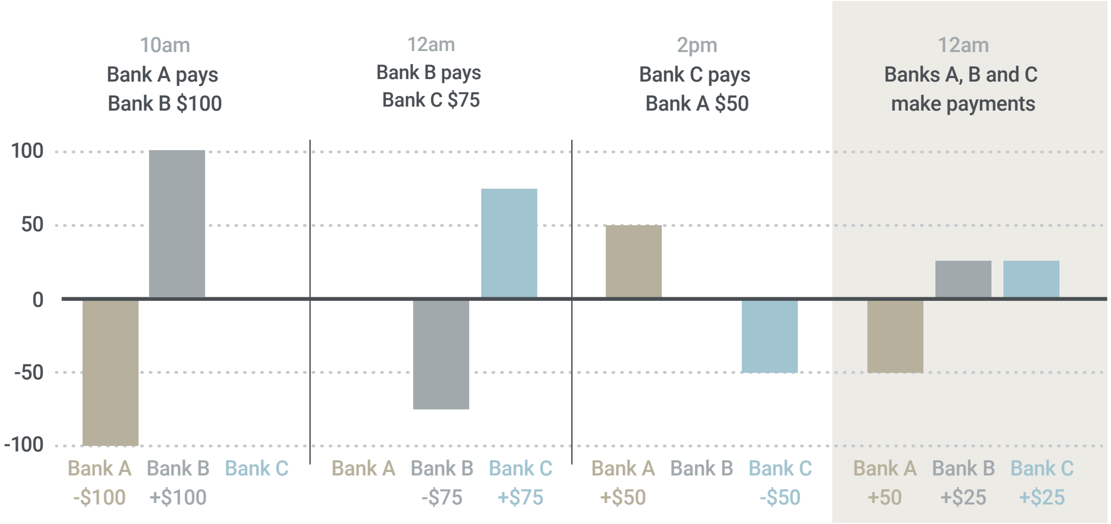
**Figure 3 – Impact of timing and coordination of intraday payments on liquidity balances**

**Asynchronous payments**

Offsetting payments made at different times result in larger changes in liquidity balances

**Synchronous payments**

Offsetting payments made at same times result in smaller changes in liquidity balances



**Source:** James McAndrews and Samira Rajan. 2000. The Timing and Funding of Fedwire Funds Transfers. Economic Policy Review, Vol. 6, No. 2. July. Available at SSRN: https://ssrn.com/abstract=888772

Although AI can be leveraged to optimize risk management processes such as modelling behavioural characteristics of non-maturity deposits and mortgage products, intraday timing of cash flows has largely remained unexplored. Moreover, Basel III requires bank participants in high value payment systems to report maximum net cumulative debit positions with a particular emphasis on payment timing throughput. The changing liquidity landscape and enhanced regulatory compliance is a strong incentive for banks to explore how AI can be used to forecast their maximum net cumulative position in order to maintain the right dollar amount of collateral without incurring unnecessary costs associated with excessive liquidity buffers. More importantly, the ability to generate what-if scenarios in the event payment timings are brought forward or delayed allows cash managers to automate payment throttling strategies that can proactively change with forecasting data or create intelligent automation that keeps decision making in their hands. The coupling of payment throttling and forecasting can potentially allow near real time update of future cash flow positioning as soon as payment timings are changed to conserve or release liquidity.

### Challenge of intraday forecasting

Direct participants in modern value payment systems that utilize RTGS with LSMs face two main challenges in developing accurate intraday cash flow forecasting.

1. Visibility of incoming payment inflow timings
2. Modelling intraday outflow payment timings to match LSM outcomes

Although sophisticated banks can make use of predictive analytics to ascertain behavioural payment timings over 30 days to one year through liquidity gap analysis and cash flow ladders, exact timing within a day is often not known as banks have different strategies in prioritizing payments as per the size, scale and complexity of their operations teams. Therefore, smaller FI participants often have reactive payment throttling strategies that mirror Prisoner’s Dilemma game theory in limiting payment outflow processing to other participants until incoming payments are received from other participants. This behaviour causes FI’s to have unnecessarily higher net cumulative debit positions over the course of a day as they become net providers of liquidity to other participants.

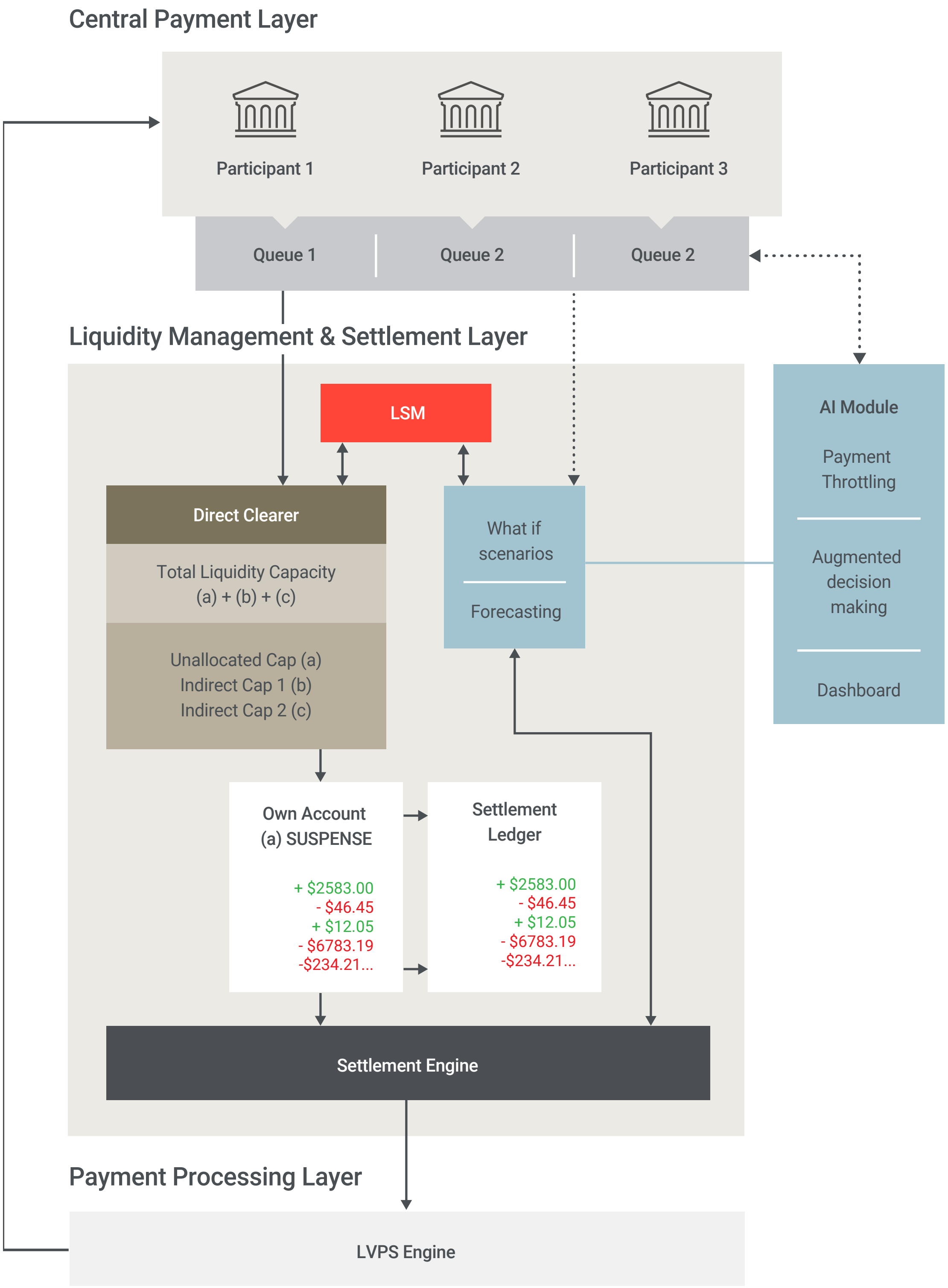
Modelling intraday outflow timings can also be complex and cumbersome if payments are processed in a non-urgent queue with offsetting algorithms. Intraday forecasting could be too conservative in assuming outgoing payments are settled immediately and some participants could encounter scenarios where forecasted net credit positions are much lower than actual positions.

## How this could work?

### Intelligent automated forecasting

Our collaborative model proposes a forecasting mechanism that leverages unsupervised learning in machine learning, performing data analysis to identify patterns without estimating a dependent variable in predicting cash flows over the course of the day.

**Figure 4: The wholesale payments system**



By interfacing directly with direct participants' internal queues and payment throttling scripts via cloud infrastructure or a payments factory that houses all payments prior to LVPS processing, the AI module aims at gaining payment visibility across all direct participants. More importantly, the AI module would interface with central market infrastructures to integrate LSM algorithm modelling in predicting the timing of payment outflow settlement from each participant. The computational power of AI is the most important consideration as predictive accuracy would largely be dependent on processing permutations and combinations of payment settlement timings.

Although unsupervised learning would be used to identify clustering of payment flows into the future, deep learning would provide choices to participants to change payment throttling strategies for optimal utilization of liquidity and payment processing. In this case, deep learning applies several layers of algorithms to incorporate LSM functionality, global market news and macroeconomic and geopolitical inputs in order to provide a series of choices for human decision making. The typical choices could be centrally administered in a user interface with simulated outcomes presented in a dashboard forecasting cash flow and liquidity balance simultaneously. For example, the AI module could identify a simulated outcome to two participants that would encourage moving the priority of payments in their internal queues forward for direct processing into a LVPS queue with LSM. The choice would have to be accepted by BOTH participants in order for the AI module to proceed. As the payments would be processed, the intraday cash flow forecasts would be automatically updated in real time simultaneously with their individual liquidity balances. Additionally, the AI would be able to generate what-if scenarios in stress testing idiosyncratic and system wide payment flows. This feature set could be crucial to provide warning indicators to participants and central market operators to avoid liquidity gridlocks.

The integration of cash flow forecasting with payment synchronization is a compelling solution to discouraging receipt reactive behaviour and minimizing liquidity usage through an unbiased central AI module. However, the centralization of outgoing payments from all participants raises a crucial concern on the robustness of privacy and security. As the payment data will not be anonymized in order to maximize analysis of client payment patterns, cybersecurity will be paramount in ensuring data confidentiality is maintained.

AI Case Study #3:

# Gridlock Avoidance in an RTGS

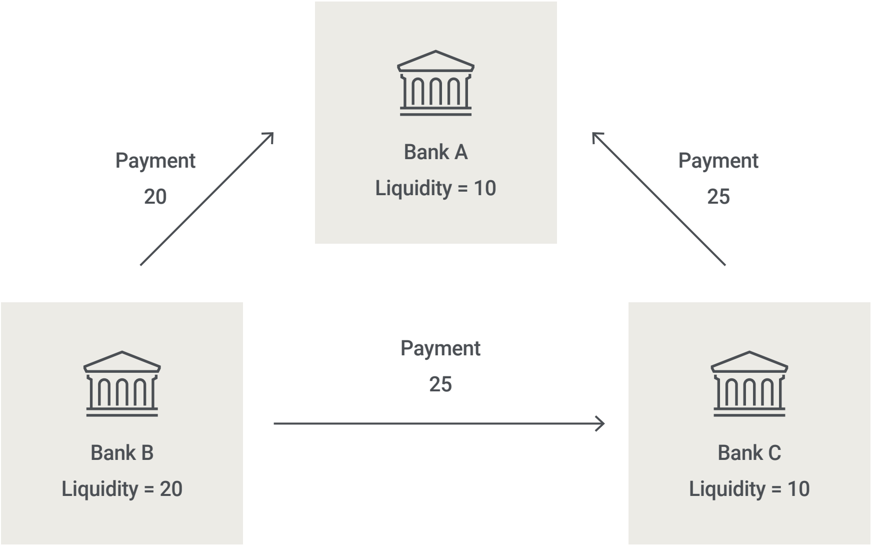
**As markets move towards true RTGS models, and liquidity challenges encourage participants to pledge only the bare minimum, the risk of gridlock in central market infrastructures rises. While there are LSMs that can be deployed to resolve gridlock, typically through simultaneous settlement or FIFO bypass, the most effective way to resolve gridlock in a market system is to ensure that it doesn't happen in the first place.**

While many market participants, as part of their overall modernization programs, are implementing enterprise-level liquidity management systems, few of these have a specific focus on core payments systems and will typically be designed to holistically support the FI rather than reflect the needs of the central marketplace. In this case, we look at the potential application of predictive analytics and AI interpretation of live data to advise participants in levels of intraday liquidity that need to be sustained on a situational basis to assist in the avoidance of market gridlock.

Gridlock typically occurs when two or more participants in an RTGS using FIFO processing block their payment queues by having a transaction at the top of their queue that is of greater value than their current available liquidity position. This essentially blocks their individual queue, and when it occurs in multiple participants can lead to a market gridlock. Perhaps more succinctly put by Bech and Soramäki[14] "*In most RTGS systems the majority of the liquidity used for settling payments comes in the form of incoming payments, and a delay in receiving these might cause liquidity problems for other banks in the system. Such formations of queues are referred to as gridlocks if the formation of queues can be attributed to the requirement for payments to be settled individually.*" Gridlock can be resolved once it has occurred through the use of purpose-built algorithms often known as 'gridlock busters' which perform simultaneous settlement on blocked queues to get the market moving again (see Figure 5). Deploying a FIFO-bypass mechanism either manually or automatically can also be used to restore the market to full function. However, both of these remedies are addressing the symptoms of RTGS gridlock, not the root cause.

[14] Gridlock Resolution in Interbank Payment Systems, Bank of Finland Working Paper No. 9/2001 – Morten Linnemann Bech, BIS Payment Systems, and Kimmo Soramäki, Bank of Finland

Figure 5: Gridlock can be resolved by simultaneous settlement



There is a fundamental truth that gridlock in an RTGS system can be avoided if participants ensure that there is a ready supply of liquidity to prevent their queue from becoming blocked in the first place. This supply can be provided directly by the participant looking to send the transaction for processing, or by a succession of inbound transactions from other participants that create liquidity for the outbound transaction. And in the same way that nuclear fusion requires an initial input of energy, so to a collateralized RTGS market requires that initial input to start the day's chain reaction of processing. However, in most countries where RTGS has been deployed, there are market accusations of 'free-riding' behaviour, best described by Martinez and Cepeda[15] "*The functioning of large-value payment systems (LVPS) can be affected when some of its participants voluntarily decide to delay their payments until they can totally fund them with the payments received from other participants. This behaviour, known as the free-rider problem, can cause an under-provision of liquidity in LVPS that operate under a RTGS mode.*"

Free-riding itself can be identified by the market regulator and measures against it including rules around payment synchronization, implementation of sync periods, managing throughput value or volume, or just simple punitive measures can be strong levers in preventing free-riding, with the overriding aim to keep the market fair for all participants and encourage positive behaviours.

With free-riding kept to a minimum, the focus for gridlock avoidance can then entail all participants collateralize to the level required for smooth and efficient operation of the market. Consequently, as shown in Figure 3 previously, a market where there are rules to drive payment synchronization, and rules governing minimum daily liquidity levels per participant can run in the most efficient way by reducing transactional delay for every transaction to near-zero and ensuring that this is done in the most liquidity efficient way.

[15] Free-riding on Liquidity in the Colombian LVPS, Banco de la República I Colombia Working Paper 977, 23 December 2016, Constanza Martínez and Freddy Cepeda.

## How could this work?

However, RTGS markets are not static systems. The daily demands on participant usage both from a volume and value perspective change dependent on market events and periodic need. This means that the 'ideal' liquidity levels for each participant and the appropriate synchronization model are difficult to monitor and effectively forecast if each participant's data is taken in isolation. Currently the leading liquidity management systems available globally such as those provided by SmartStream, TAS and Broadridge, are fed from the data of an individual participant—in the case of payments, the incoming and outbound payments made by the bank. This means that liquidity optimization becomes a competitive area which will generally result in behaviours that are negative to the fair and efficient running of the wholesale payments market, with the outcome typically being gridlock. Assuming that a fair and efficient market are the goals being strived for, then leveraging predictive analytics and AI that has access to whole market data, i.e., driven from the centre, the market can be informed of the criteria needed to create the best conditions for smooth operations and there is a net benefit for all participants.

Looking at how this could be enacted is relatively simple. By creating a market model based on historical data, analysis could be performed to calculate for any specific historical day what the minimum level of liquidity would have been for every participant to process all of their transactions without delay. This would be the baseline for the level of liquidity that each participant would need if the same day were to be repeated. The model could then further be enhanced to identify transactional trends and recent liquidity usage to further optimize this level and so forecast the required minimum liquidity needs for any participant on any given day. The supervisory body could then mandate this as the minimum threshold for daily participation.

Furthermore, intelligence built into the system could help to make it more proactive from an intraday perspective. Should the liquidity position for a participant be lower than expected at a certain time of day due to either an unexpected market imbalance or increase in outbound transactions, a system could be built to warn the participant that they are running below predicted liquidity levels, and knowing what is likely to come, suggest remedial action through intraday liquidity injections or other market methods to improve their liquidity position.

Another interesting manner in which AI can help central market infrastructures mitigate payment gridlocks could be automated decision making to move liquidity balances between pure RTGS queues to LSM queues for non-urgent payments. The key differentiator would be the proactive movement of balances by a central market AI module without the involvement of participants, as central market infrastructure's visibility of all payment flows would make it better placed to forecast payment gridlocks. The continuous movement of balances between these queues would optimize liquidity efficiency ratios and also help free up participant personnel time from operationally intensive tasks in a real time environment. By anticipating the market and self-regulating participants' collateralization, there can be significant savings due to the overall lower liquidity needed to ensure an efficient RTGS. More importantly, this centralized feature could help in a stress event, where AI regulation of collateral management would restrict idiosyncratic behavior emanating from game theory and help streamline payment processing.

AI Case Study #4:

# Cost-driven Market Smoothing

## One of the major changes to the functioning of markets in the app-era has been the application of predictive analytics and artificial intelligence to aid in their efficiency.

One of the clearest examples of this is the surge pricing employed by ride-hailing apps Uber and Lyft. Controlling a market through variable or punitive pricing is not a new idea, but as we move towards the next generation of AI and have exabytes of data covering 20+ years from which to build predictive models, the ability to control markets to increase efficiency has taken a massive step forward.

As indicated in the previous examples, markets with volatile transactional traffic can be inefficient both from a processing perspective and from the inability to match both sides of the market. As explored in the previous example, this volatility could lead to shortages in liquidity that lead to processing delays or even gridlock within the market, both of which can have a negative impact on the users of the system. While the ride-sharing market uses surge pricing to ensure that supply of drivers matches demand of customers waiting to ride, the same mechanism can be re-tasked to encourage behavioral change that will improve overall efficiency.

Synchronization, which is the spread of transactions throughout the day in a liquidity-backed settlement system, has been long recognized as the most efficient liquidity savings mechanism. In 1999, a Bank of Finland research paper[16] elucidated this point, "*In a real-time environment banks face new challenges in liquidity management. They need to plan for intraday as well as interday fluctuations in liquidity. Not all payments in a real-time environment require immediate processing. This gives system participants the opportunity to employ different types of hybrid settlement structures, which enables the evening out of intraday fluctuations in liquidity demand.*" However, enforcing this behavior without adequate policing and strong punitive measures is typically beyond the resources of most central market infrastructure operators. Typically, the participants have far more bodies thinking about how to 'game' the system than the payment market operators have attempting to keep it running fairly.

[16] Optimizing Liquidity Usage and Settlement Speed in Payment Systems, Bank of Finland, Harry Leinonen – Kimmo Soramäki, Financial Markets Department 12.11.1999

Although some of the payments passing through a settlement system are absolute in their time dependency, for example payments to the central bank or settlement of securities obligations, there are many transactions that pass through these platforms that require only same-day settlement. A mechanism to encourage better distribution of these non-urgent transactions could have a significant impact on the liquidity efficiency of both the participants and the market as a whole.

## How could this work?

Leveraging the significant legacy data available for central markets and predictive analytics, a daily pricing schedule could be created to discourage users from sending non-urgent transactions at points where there are significant movement of known time-dependent transactions. The pricing could also be used to encourage sending of these transactions at low volume times for the market, smoothing out overall liquidity usage and creating better payment synchronization with the reciprocal increase in overall market efficiency.

However, surge pricing is only efficient as a market measure if it drives the preferred behavior. If a transaction with a processing cost of $15 can be sold to the participant's client for $60, raising the processing cost to $16 will likely not have a large impact on FI behavior as it will be grudgingly absorbed or easily passed on. If the processing cost were to be raised to $60—equal to the external market cost—this may become a disincentive to processing at that precise moment and have an impact on transaction timing. However, experimentation would need to be done to find the pinch point at which surge pricing begins to reduce overall market traffic and potentially create parallel markets or drive end users to other potentially less-secure transactional systems.

It is expected that over time the necessity to deploy surge pricing would be cyclic in nature. As pricing initially bites, it would drive positive behavior, but as the market then responded with lower surges, the market could snap back and so a cycle would endure.

## Open Questions

There are times where operational failures for a market participant may leave them unable to maximize efficiency across an entire period and forced to process an entire day in a compressed time frame[17]. It could be seen that their need to 'eat' increase processing fees for failing to invest in their own resiliency is appropriate punishment for the disruption to overall market efficiency but could in the case of a less-liquid participant drive them towards default, which wouldn't be in any party's interest.

Furthermore, if the market is operated on a cost-recovery basis rather than a pure commercial basis, there is a strong case for fee reduction at 'quiet' times as well as the reciprocal surges to pricing.

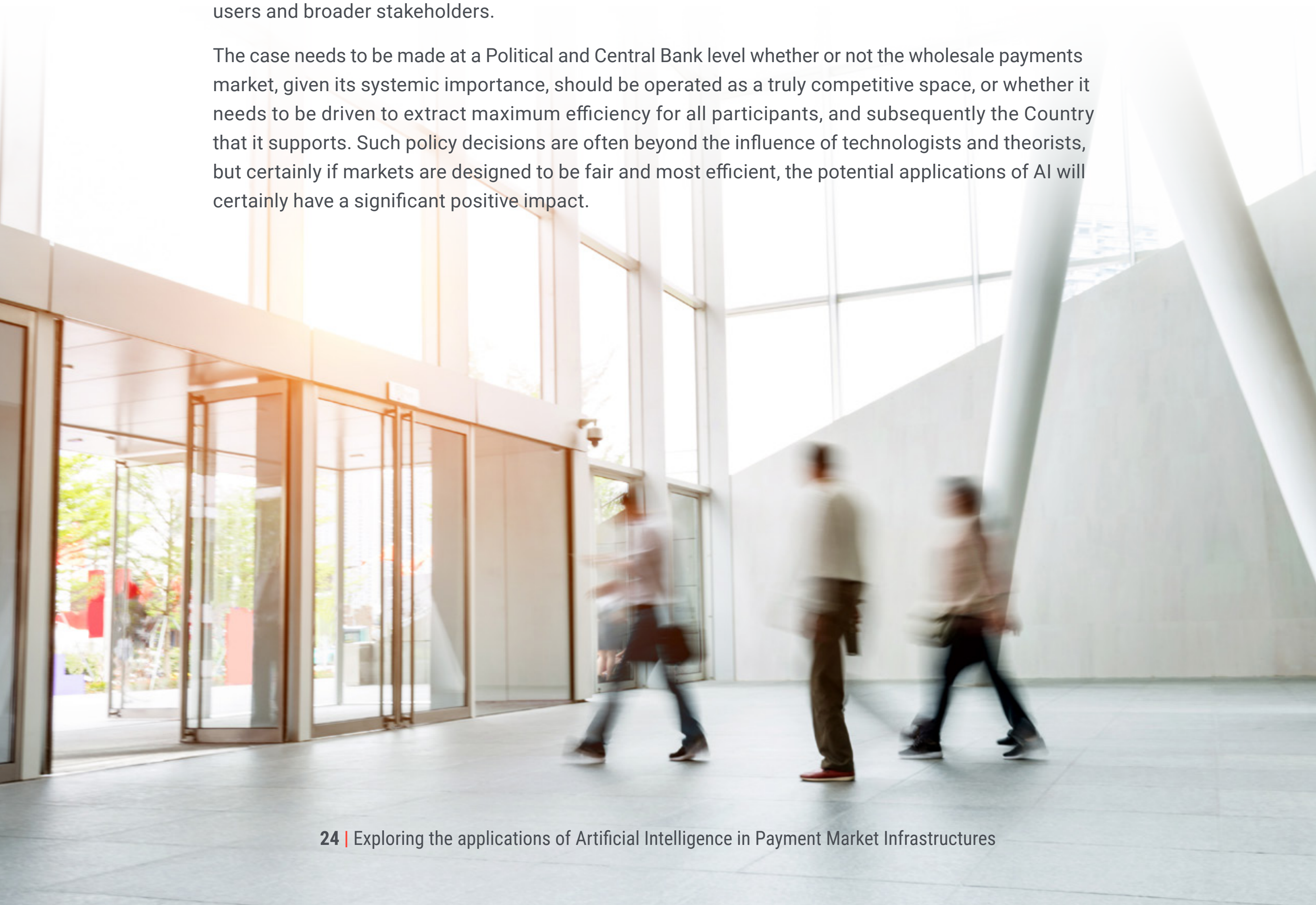
[17] For example, the IT failures suffered by the UK's TSB Bank in 2018 which prevented processing for a 48-hour period

# Concluding remarks

It is clear that wholesale payment markets can be made significantly more efficient through the implementation of one or more of the described use cases. However, the technological underpinnings of many of these systems are still aligned to more traditional stacks making implementation more difficult. As these systems are redesigned and replaced, such as is happening in Canada today, there is a significant opportunity to evaluate placement of AI tools into the future roadmap and take these infrastructures into the future.

As mentioned, given the critical importance of these systems to national financial and economic well-being, it is also imperative that implementation of this technology is carried out in compliance with regulatory direction around system safety, security and information privacy to protect the interests of users and broader stakeholders.

The case needs to be made at a Political and Central Bank level whether or not the wholesale payments market, given its systemic importance, should be operated as a truly competitive space, or whether it needs to be driven to extract maximum efficiency for all participants, and subsequently the Country that it supports. Such policy decisions are often beyond the influence of technologists and theorists, but certainly if markets are designed to be fair and most efficient, the potential applications of AI will certainly have a significant positive impact.



CGI

Founded in 1976, CGI is among the largest independent IT and business consulting services firms in the world. With 77,000 consultants and professionals across the globe, CGI delivers an end-to-end portfolio of capabilities, from IT and business consulting to systems integration, outsourcing services and intellectual property solutions. CGI works with clients through a local relationship model complemented by a global delivery network that helps clients digitally transform their organizations and accelerate results. With annual revenue of C$10.8 billion, CGI shares are listed on the TSX (GIB.A) and the NYSE (GIB).

**Learn more at:** cgi.com/canada
or info@cgi.com

PAYMENTS
CANADA

Payments Canada ensures that financial transactions in Canada are carried out safely and securely each day. The organization underpins the Canadian financial system and economy by owning and operating Canada's payment clearing and settlement infrastructure, including associated systems, bylaws, rules and standards. The value of payments cleared and settled by Payments Canada in 2018 was $53 trillion or $209.7 billion each business day. These encompass a wide range of payments made by Canadians and businesses involving inter-bank transactions, including those made with debit cards, preauthorized debits, direct deposits, bill payments, wire payments and cheques. Payments Canada is a proud supporter of the Catalyst Accord and the 30% Club.

**For more information about Payments Canada, please visit:** payments.ca