WHITE PAPER

Digital Risk Management & AI: Mining Risk from Disparate Data

Contents

Introduction

In today's global economy, businesses' technology portfolios are increasingly complex, varied, layered, accretive, and expensive to acquire and maintain. Most businesses are becoming digital, even those who produce physical, retail goods and maintain a brick-and-mortar presence. On the flip side, many companies born online are establishing physical facilities, complicating their technology portfolios. This introduces the added complexity of public-facing interfaces, vendor-credentialed or open APIs, and content delivery networks (CDNs) in multiple global locations with cache.

These factors are compounded by external forces such as new regulations around privacy, the press for environmental, social and governance (ESG) compliance, and increasing threats of cybercrime. Add in the proliferation of professional productivity platforms like Microsoft Office, its competitors, team collaboration, bring-your-own-device (BYOD) policies, email, file stores, virtual conferencing, hosting, code repositories and anything else that takes an attachment or upload. On the other side, the majority of customers are now fully immersed in the digital world, and expect their trade to be accommodated anywhere, everywhere, and anytime.

Underneath it all, there is a plethora of disparate data sources and infrastructure, including hybrid hosting. While AI is generally incapable of human heuristics, its statistical reasoning capabilities makes it good at discovery in mountains of data. In this paper, we will focus on how AI can be used to identify risk from low-level data layers. We will also introduce some processes to manage the jetsam and funnel leads into an operational queue for CISO, CCO, IT and security teams to investigate.

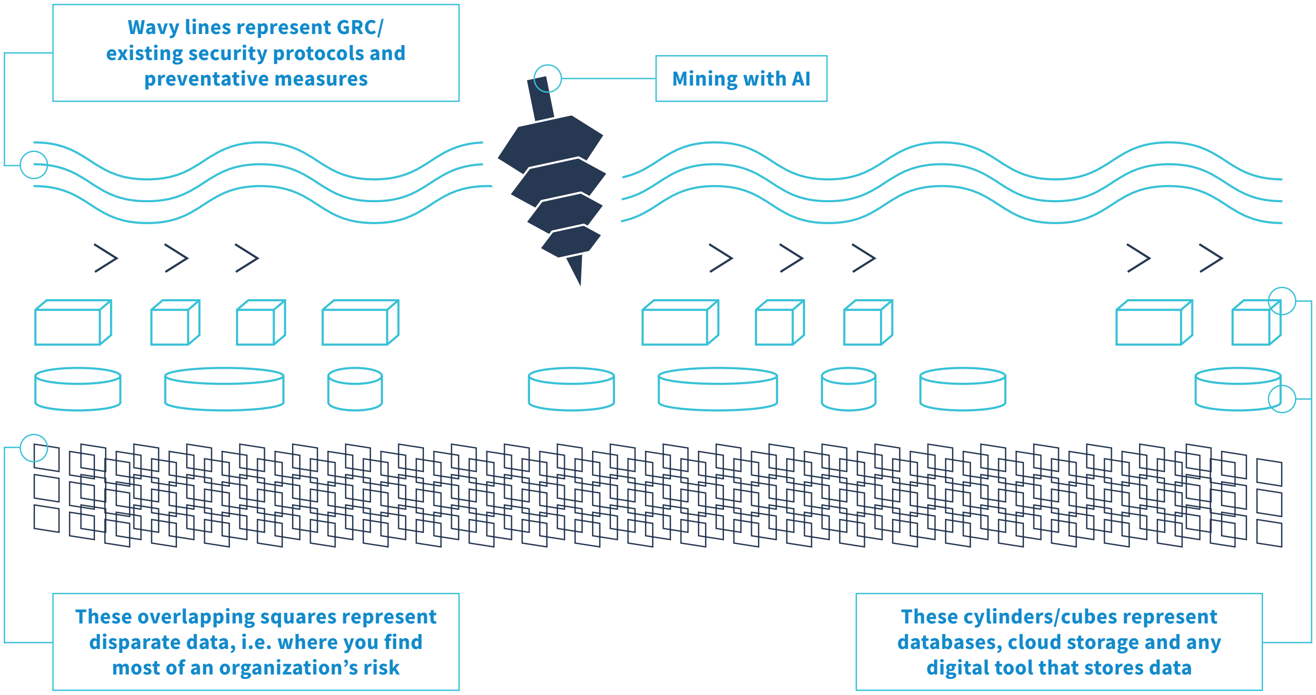

Wavy lines represent GRC/ existing security protocols and preventative measures
Mining with AI
These overlapping squares represent disparate data, i.e. where you find most of an organization's risk
These cylinders/cubes represent databases, cloud storage and any digital tool that stores data

Figure 1: *Mining for risk from disparate data layers within an organization's digital repositories*

THE AI TECHNOLOGY STACK: CONSIDER THE CONTEXT

If there's one thing we've learned over the years, it's that AI depends wholly on immediate context. Consider the types of risk that could have the most impact on your business. Is it something related to privacy, competitive intelligence, intellectual property, reputational risk, criminal vulnerability or a security breach?

Before mining for risk with AI, companies need to take inventory of the technology they have at their disposal. There are a range of digital tools that focus on various aspects of risk. The power of these tools lies in their ability to unearth data outliers and expose potential risks. Part of the evaluation should be to understand each one's strength in the domain they are targeting, and whether they complement, improve or supplant in-house DRM resources.

Depending on the industry or use case, existing software can be applied. In the case of out-of-the-box, pre-trained models, align the trained domains with your roadmap, with scale as an added consideration. Some generalized solutions seem miraculous during a demo, but don't always translate to specific business functions.

Risk mining with AI is an emerging application, so new products on the market are still maturing. It's fair to expect some volatility in the evaluation process. There are numerous companies focused on this space, along with incumbents releasing new products or refactoring existing tools.

It's critical that the solution can connect to the full array of storage and access the entire network. By focusing most of its information processing techniques on risk frameworks, rather than other use cases, machine learning (ML) becomes highly targeted and more effective.

Once selected, the software should be run in multiple modes to help it fill the gaps that result from tight, articulated models, especially the well-designed ones. This complementary deployment of raw statistical discovery will find more random associations and cluster artifacts or database tables that contain potential risks.

Introduction (Cont.)

OUTSIDE EXPERTISE: CONSIDER COMPLEMENTARY CAPABILITIES

It may help to bring in expertise from a vendor to complement in-house capabilities, or at least advise in-house experts. If selecting a vendor for collaboration, consider the following key criteria that may help optimize the investment:

- **A dedicated governance, risk and compliance (GRC), and DRM practice**
- **A CISO with robust enterprise capability and substantial consulting capacity**
- **Global reach with presences in most major jurisdictions that your company operates in**
- **Broad and deep engineering experience**
- **Familiarity with regulatory operations in multiple industries and jurisdictions, with commercial legal domain expertise**
- **Experience assisting clients with change and transition due to corporate actions, significant reorganizations, and business or product rationalization**
- **Industry-specific business consulting capabilities**
- **Advisory capabilities on the technology process**

The upfront costs of data mining can be significant, both to start and as an ongoing operation. However, when executed well, it can prove to be a worthy investment, as it offsets future expenses that would arise from a security breach or regulatory non-compliance.

It's important to have meaningful executive support and governance to cut through complex cross-functional, technology and operational issues, as well as honoring controls, while securely removing barriers to enable the initiative. With a roadmap in place, tools selected, staffing ready and processes defined, mining can begin.

Managing the Mining Process

This is the time to mix in traditional DRM techniques. Mining can generate significant volumes of output that require a rigorous and disciplined mitigation and resolution effort, which will start alongside the baseline initiative.

Once past stage-one in priority domains or throughout the enterprise, it's time to transition into more of an operational mode. To begin, the initial passes can generate a baseline. This gives additional input to statistical reasoning to identify repeating risks, or outliers that need investigation.

This simple and powerful technique is not without complexity—mining brings its own complications. To be specific:

- **Be careful of the detritus that data mining leaves behind to avoid creating new risk**
- **Picking a starting point can lead to paralysis by analysis—don't overthink where to begin, follow the logical course**
- **Be sure to line up the expertise to traverse wide variations of process, technology and organizational units**
- **Data mining is not likely to be a one-time exercise; this is a capability that gets optimized and includes its own trail of active monitoring technology bits**
- **Don't underestimate the training and accuracy cycles needed for ML, it can be as costly as good old-fashioned artificial intelligence (GOFAI) coding**

If possible, start off with two or more of these or similar domains. Try to vary size, complexity and types of risk, as well as common and outlier cases. Providing early results will help the core team, stakeholders and governance pick up velocity, while teasing out specifics in a more complex domain that will make work processes more robust. Also try to get a mix of structured and unstructured content to assure breadth of capability.

The good thing about using this highly targeted ML approach, complemented by the rest of your DRM program, is that the fuzzy edges typical of accuracy around ML don't matter. This strategy should trigger defensive action rather than exactitude and purity of processing.

Don't underestimate the effort and expertise needed to evolve, finesse and diverge the training sets for the ML model catalogue that's plugged into the mining engine. However, don't try to plan for all the training in advance. Use an agile approach that enables discovery, optimization or scale as needed. Give the AI some basic training in identifying and associating the data that highlights risks.

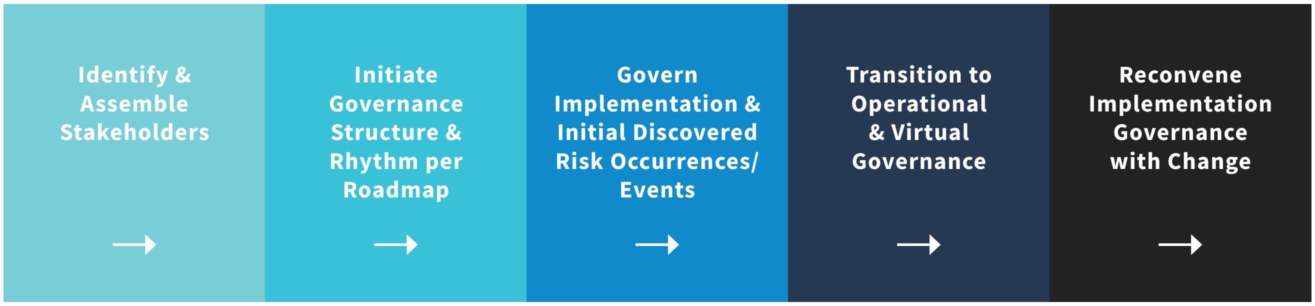
GOVERNING THE BASELINE INITIATIVE

The initial effort to tunnel across the enterprise and establish a baseline will necessitate engagement with at least two types of governance, business ownership of data and IT/audit controls around it. An important decision will be where to situate the governance in the organization. Three obvious choices include:

- **Independent.** Usually this is the route chosen if the effort is to be confidential or closely quartered. However, there is a risk of governance fatigue, and it's likely that those who would be in this risk mining initiative are likely in one or more other governance groups. We would default to combining governance with one of two other related cross-functional governance organizations.
- **Alongside data governance.** Data governance is the most likely choice due to roughly the same terrain. This means most issues are well understood by the group already, and the necessary gating and consent processes can be reused.
- **Alongside security governance.** Security organizations will likely be a high priority stakeholder for this activity. If less visibility is desired, for example in an organization that has a large staff, structured hierarchically with a wide lower rung of employees should be employed (versus a smaller, more networked organization of mostly mid-level professionals or above).

The most significant impact for risk mining will be working through typical data control topics like privacy and security, and avoiding loss or exposure of confidential information. After an initial landscape discovery to work through, along other key data topics, a comprehensive roadmap can be developed.

One last topic for consideration is the level of monitoring, supervision and an audit trail from trawling. Consult with in-house counsel and outside legal and regulatory experts about what to document, how much to document, who'll be able to see it and the proper auditability. This is not necessarily a case of more being better; in some cases the minimum is best.


Identify & Assemble Stakeholders →

Initiate Governance Structure & Rhythm per Roadmap →

Govern Implementation & Initial Discovered Risk Occurrences/ Events →

Transition to Operational & Virtual Governance →

Reconvene Implementation Governance with Change →

Figure 3: *Governing the Baseline Initiative*

Managing the Mining Process (Cont.)

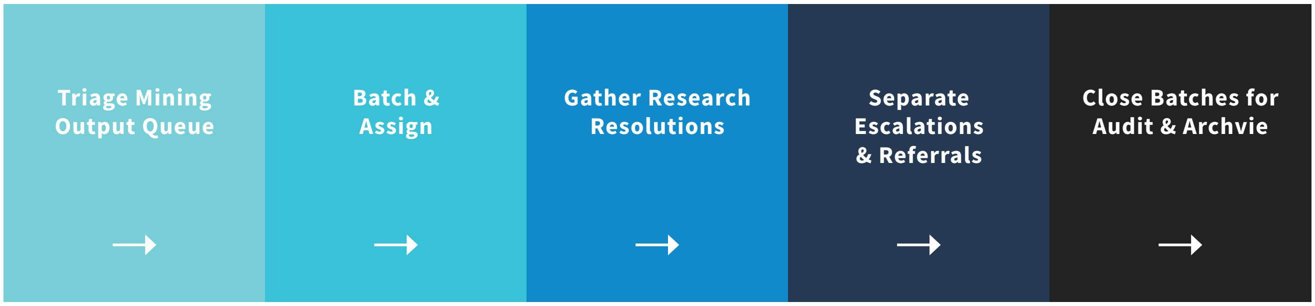
TRIAGE, BATCHING, ASSIGN & CLOSE

The first step is to tune the output, whether via scoring or expected workload, to get batches that fit into an agile sprint of one to three weeks. Each batch should be prioritized, most obviously by weights, scores and distributions from the ML output. Plan the timing to coincide with governance rhythm—any serious issues can and should be escalated immediately.

Prepare for anomalies by having a reasonable cadre of analysts who investigate anomalies on standby. Look for a team with a wide variety of expertise ranging from language to functional knowledge and technical knowledge.

The bulk of the work will be chasing down edge cases of regular business processes, or minor functional violations, like an accounting slip-up. There might also be the occasional ML hiccup, like a series of unrelated words, that when statistically glued together, create a red flag. This would be the case for an email that reads, “That last release was the bomb. Do you think it will slay the Vice President of the division?”

Additionally, data elements can help, but be careful of weighting a million dollar anomaly over a missing penny. On one hand, the penny may be two offsetting anomalies, or nefariously planned to seem unimportant on purpose.



Triage Mining Output Queue →

Batch & Assign →

Gather Research Resolutions →

Separate Escalations & Referrals →

Close Batches for Audit & Archvie →

Figure 4: *Triage, Batch, Assign & Close*

Managing the Mining Process (Cont.)

INVESTIGATION, RESEARCH & RESOLUTION

This involves setting standards and operating procedures for investigations related to documentation, research, recommendations, resolution and supervisory review. The approach for analysts should be collaborative and agile; let them help each other. The ability to dialogue helps eliminate oversights and sharpen insights, which contributes to producing better results. Both the weighting and scoring factors should be reviewed for tuning with each batch, as well as at milestones like the first pass through a domain.

Resolutions may vary. With research, innocuous anomalies that are a part of normal business can be closed by the analyst, in most situations. A supervisor or approver may want to review or explore randomly from a queue with summary information. The analyst can initiate some mitigation or remediation, follow up with the assigned and close after remediation. These should be documented with the anomaly.

In many cases, mitigation and remediation may require an extended workflow or escalation. There may be existing queues for analysts to refer issue into other functions. Most institutions have numerous security anomaly workflows and operational procedures.

Early on, most upward escalation is offline, unless your situation generates a high volume of follow-up, such as in securities or consumer insurance. You may add on or closely couple escalation to your analyst workbench in later phases, or as otherwise required.

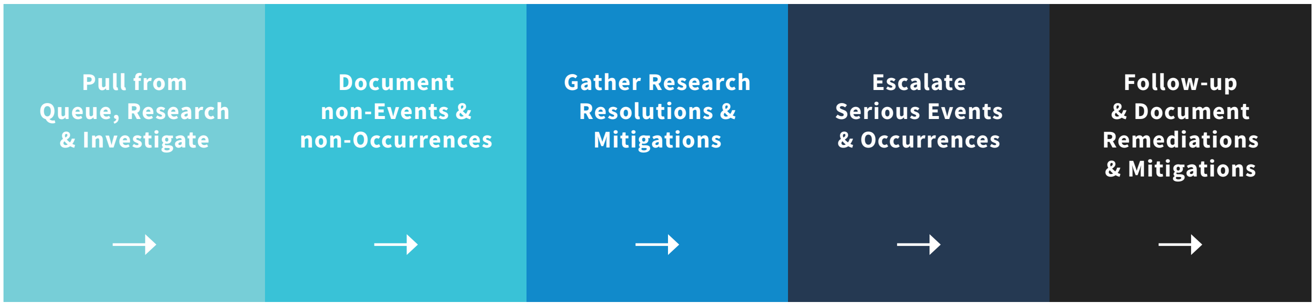

Pull from Queue, Research & Investigate
→
Document non-Events & non-Occurrences
→
Gather Research Resolutions & Mitigations
→
Escalate Serious Events & Occurrences
→
Follow-up & Document Remediations & Mitigations
→

Figure 5: *Investigation, Research and Resolution*

Analyst-Initiated Investigations

In some cases, analysts may initiate investigations. On their own, analysts should be able to start an investigation for any reason. Typically, these may be to fill up slack time, routine operating procedures for slack time or periodic triggers such as closes. Also, in some industries and functions, employees or external entities may request investigation. Examples include tip lines, whistleblowers, and customer inquiries over an oddity in their account activity.

In other cases, analysts may want to review mining logs and output in an exploratory fashion. AI does not know what it does not know. A person with heuristic capability may spot patterns that have not yet been trained into the AI or are beyond the AI's capabilities. In these instances, platforms with an analyst workbench should allow analysts to add to the queue in various ways, whether or not originating work items initiate from the mining engine's output.

Organizations may have existing risk procedures that should be consolidated into the analyst workbench. Consider that current COTS risk mining suites, if they include an analyst work bench, may not be configured for externally initiated cases added to the queue, and require customization or integration.

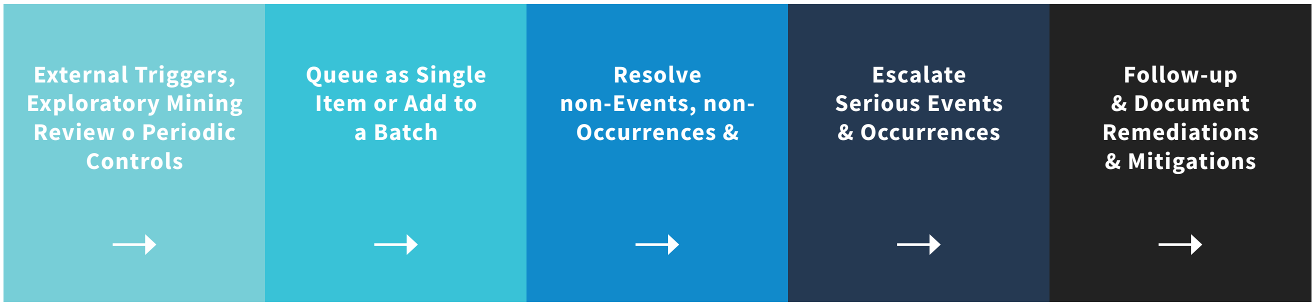
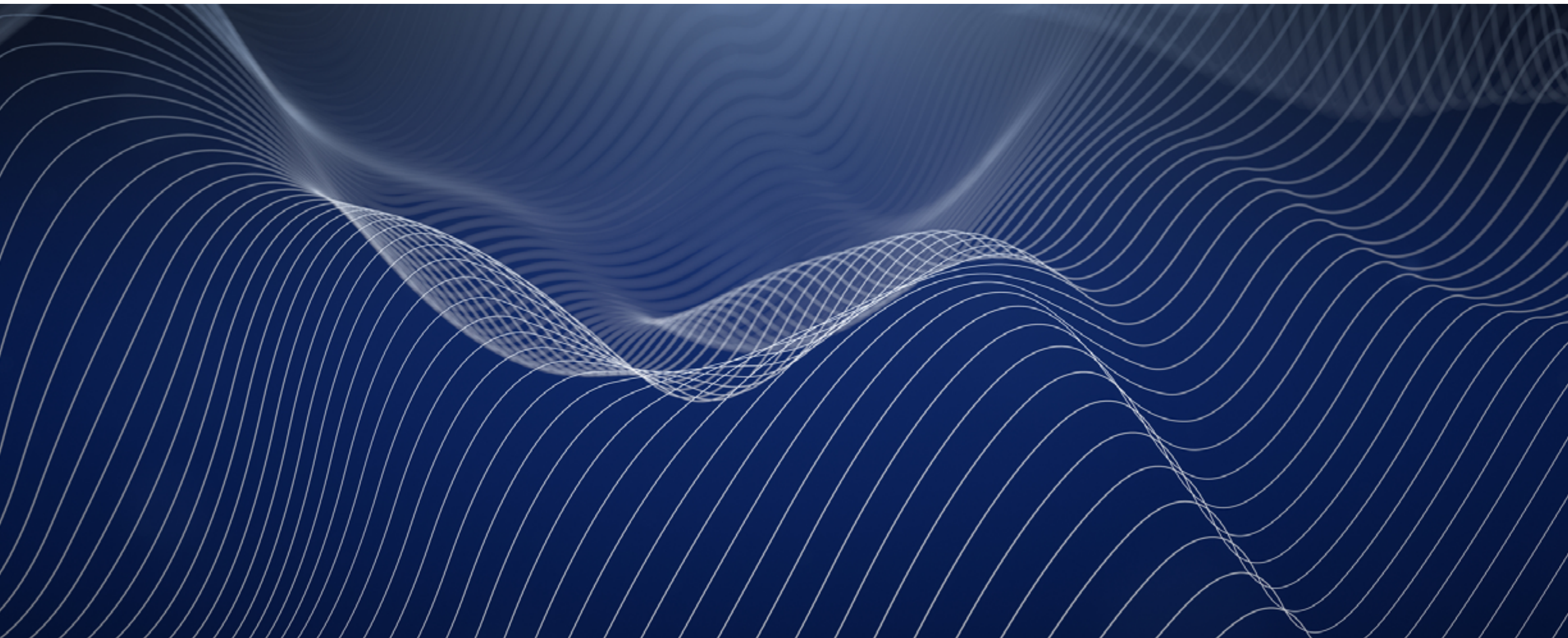

External Triggers, Exploratory Mining Review o Periodic Controls
Queue as Single Item or Add to a Batch
Resolve non-Events, non-Occurrences &
Escalate Serious Events & Occurrences
Follow-up & Document Remediations & Mitigations

Figure 6: *Analyst Initiated Investigations*



Model Accuracy, Optimization & Supervision

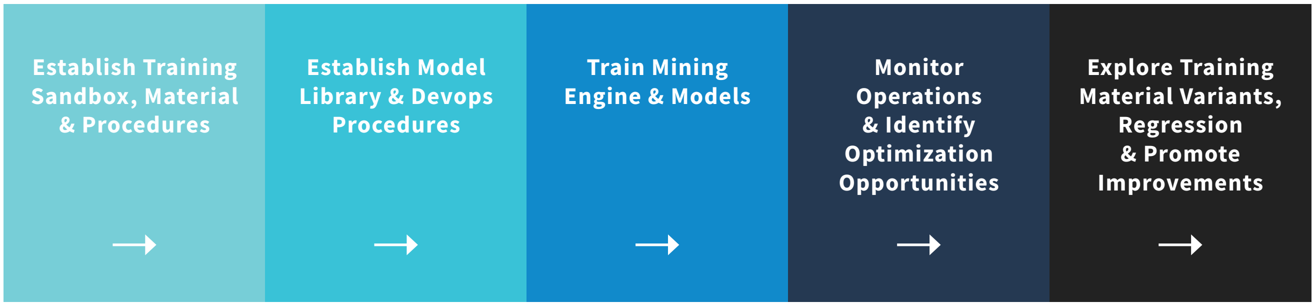
Occasionally, it becomes apparent that institutions have an innate bias that underestimates the work of teaching AI what it needs to know. Planners sometimes assume that ML requires less cost for training models than GOFAI expert-based options with rules, regular expressions, vocabularies, taxonomy, ontology and so on. Nothing could be further from the truth. You will need to account for AI upkeep and validation in initial phases and into the future. Considering this maintenance, budget should be carefully delegated.

Recently published articles associated with a well-regarded chatbot described a staff of 90 working on a bot. These articles did not report initial ROI forecast versus actual results. GOFAI expert ontology-driven bots tend to require significantly fewer resources. That said, the ML-driven mining we suggest here has advantages over expert mining. It may detect risk that is beyond the scope of your institutions expert knowledge altogether, let alone the expertise codified as rules in a GOFAI engine. Our recommendation is to put aside the hype and get real information on the costs of training ML for this purpose.

While optimization can be a long tail, change in contemporary business and markets is accelerating. Impact from change will force organizations into re-establishing initial phases to implement across the change domain. Consider a simple acquisition of a small company—if any of their IT assets remain, it's imperative to extend DRM mining to cover them. With larger acquisition, this effort and cost can be significant.

In some cases, the acquisition may be doing something similar, though there should be some integration and common ways of working across both portfolios, which can mean choosing the better platform and refactoring or integrating the two. This is especially important if the acquisition was complementary and not accretive because that acquired risk mining platform is better suited to the new territory.

Consider that in a divestiture, you may want to wipe your AI risk mining from the assets you hand to the buyer. There is no point in revealing your mitigation capability beyond the front door. Your buyer also may not have the discipline you do, and penetration or failures of their DRM can hand over the keys to your institution to nefarious entities.



Establish Training Sandbox, Material & Procedures →

Establish Model Library & Devops Procedures →

Train Mining Engine & Models →

Monitor Operations & Identify Optimization Opportunities →

Explore Training Material Variants, Regression & Promote Improvements →

Figure 7: *Model Accuracy, Optimization & Supervision*

Ongoing Governance, Reporting & Auditability

After going through the initial implementation, you will want to transition to a lighter, ongoing governance focused more on the operational activities associated with the mining results. In most cases, this will resemble traditional controls such as finance, fraud and market impacts. If an organization has enough to have a low operational workload, that does not mean the governance group is off the hook.

Be sure to include periodic drills of various types, and stress testing the capability. Also consider engaging a partner that specializes in introducing risk on a trial basis, such as penetration tests, virtual customer, virtual vendor or other simulation actions, etc. In purely operational mode, consider giving back valuable stakeholder time by shifting to a virtual governance structure, rhythm and connection. Living dashboards or a stakeholder workbench can facilitate this.

There are frameworks and software for this virtual governance, including the ability to convene more fully in real time as needed. One of the virtues of virtual governance is the ability to act or react more quickly. This capability may come in handy for risk occurrences and events that turn up unexpectedly as this new capability is applied across the business portfolio.

Conclusion

Using AI, and especially ML, to crawl your low-level enterprise data can be a powerful tool in protecting the entire company from once hidden risk. By maintaining a cautious, investigative approach and carefully selecting the right tools and analysts, you are more likely to see the desired results without creating unnecessary organizational impact. View the budget not as an expense, but as cost reduction for future expenses avoided, and estimate the cost of future risk events when calculating ROI. Plan for robust governance to clear the way, and deal with sensitive escalations.

As part of the consideration for what happens after the mining reveals anomalies, plan how to transition from the initial phase to ongoing operating procedures, and deal with the types and volumes of risk events surfaced by the mining. Collaborate with trusted advisors that can bring objective expertise from the broader DRM community, and other industries. Most importantly, maintain an agile approach and adaptable response when anomalies are discovered. This will help to set protocols in place that organizations can easily replicate and ensures future risk mining is successful.