FBED LABS. / RESEARCH

Research Note 001 • 6 November 2026

The Physics of a Sale

How commercial decisions become observable, testable and learnable.

By **Joshua Bardsley** / Founder & CEO, FBED Labs

The question beneath the words

A buyer says, “This is interesting.”

The seller leaves the call encouraged. There was a good conversation, a real problem perhaps, the beginnings of a relationship. The CRM is updated, the notes are tidy, and the opportunity looks promising.

A week later: “They're coming back to me.”

A few weeks after that: “We're talking again next month.”

And then, very often, nothing.

I have watched this happen across complex commercial transactions more times than I care to count. Sometimes the buyer returns. Sometimes they were never going to buy. Often there was a moment when the opportunity changed, but by the time we recognise it we have lost the evidence required to understand why.

As the need for a result grows, the volume of action tends to grow with it. More calls, more messages, more follow-ups, more forecasts. That activity creates signals, and the same pressure that drives it can colour how we read them. A polite answer begins to feel like interest. Interest becomes intent. Intent finds its way into a pipeline as expected revenue. We start treating the fact that somebody *might* do something as evidence that they *will*.

Hope is not a bad thing. A company needs it. A seller needs it. It is a terrible substitute for measurement.

I sometimes think of conventional commercial systems as **hope machines**. Not because the people building them are unserious, or because modern tools cannot identify authority, intent, objections and next steps. Many can, and increasingly well. The deeper problem is that much of this intelligence is assembled *after* an interaction. It records what was said and what the seller now believes. It does not necessarily preserve what was known *before* the decision, which alternatives were available, why one was selected, what was actually delivered, and what would have to be true before we could say it helped.

A summary is not a decision record. A retrospective explanation is not causal evidence. A plausible next step is not a measured improvement.

Artificial intelligence has become astonishingly good at language. It can role-play a buyer, write a strategy, compress a transcript and explain a lost deal. But a model that sounds like a seller does not necessarily know how a sale changes, and an articulate simulated buyer is not proof of how a real

buyer would have behaved. Improvements in general model capability are not, by themselves, evidence of improved commercial outcomes.

The part that interests me is what might exist beneath the words.

I believe there is order to commercial behaviour on a level we can learn. Not certainty, not a universal script, and certainly not control over another person. I mean patterns in how people respond to information, weigh risk, grant trust, defer, make commitments and change their minds. Patterns conditioned on circumstances we only partly observe, played out across decisions whose consequences may take months to become visible.

If that order exists, we should be able to investigate it. We should be able to say what we knew at the time, what we tried, what happened next and—when the evidence supports it—whether a different decision might have produced a better outcome.

That is the question beneath Shannon, our research programme at FBED Labs. It is why we started with commercial state, decisions and observed trajectories rather than another layer of generated advice. The aim is not to control buyers. It is to understand where our beliefs diverge from reality, identify moments that may be worth intervening in, and give people and systems a better basis for their next decision.

This paper starts with the shape of a sale, passes through statistics and reinforcement learning, and ends with Shannon Zero: a model we intend to build and test against the evidence that precedes it. Every important scientific idea has a proper name. I will introduce each one as we need it, because I think the reader deserves to understand the problem before being asked to admire the mathematics.

1. A game nobody has written down

A sale is a journey. A person or company begins somewhere: a reply to a message, an introduction, a first conversation. There may be discovery, another conversation, an offer, procurement, silence, a decision. The seller acts, the buyer responds, and the world changes between their interactions.

I was once asked whether this was like chess. My answer was no, although chess provides a useful way to begin.

In chess, each player can see every piece. The rules are fixed, and one player's victory is the other's defeat. A healthy commercial transaction is different. Both parties can gain, neither can see the other's complete position, and the conditions under which they are deciding may change while nobody is speaking.

A seller does not know everything the buyer knows. They may not know how serious the problem is, who else can veto the decision, whether there is budget, what happened in the last internal meeting, or even whether their counterpart trusts them. The buyer has an equally incomplete picture of the seller's capabilities and intentions.

But this does not make a sale structureless.

Nor is a complex sale a simple battle of persuasion. A buyer may want the solution and still be unable to proceed; the seller may want the transaction and still be wrong about whether it would help. Their interests can align, conflict, or change as the decision moves from one person to a committee. There are moments of cooperation and moments of adversarial negotiation. There are also delays and decisions that neither side controls.

If behaviour were perfectly predictable, we would not need a model that learns. If it were entirely without regularity, a model would have nothing to learn. The interesting territory is between those two extremes.

One way to model sequential decisions is a **Markov decision process**, or **MDP**. At its simplest, it asks us to specify four things: the *state* of the world, the *actions* available, the *transitions* that might follow an action, and the *reward* by which progress will be judged. This is a modelling framework, not a claim that every real sale already satisfies the mathematical assumptions of an MDP. [1]

Once the decision-maker cannot observe the full state, a related framework becomes useful: the **partially observable Markov decision process**, or **POMDP**. Instead of knowing the buyer's true position, we hold a belief about it, update that belief as evidence arrives, and choose actions under uncertainty. [2]

Those names may sound remote from a conversation. They describe something that a capable seller already does instinctively.

They enter with a picture of the buyer. They listen for evidence that changes it. They choose what to ask or say next. They revise their plan. The important difference is that a learning system cannot rely on an expert's unrecorded intuition. It needs a representation that can be inspected, challenged and, eventually, tested.

Our starting point was therefore not the policy that chooses the next move. It was the harder question: *what would the policy need to know?*

2. Seeing the board

Before a good seller makes a call, they are carrying a model of the room.

Is this company growing? Is the problem painful or merely interesting? Does the person have authority? What evidence would they need before believing us? Which parts of our understanding are solid, and which are guesses?

There are two ways to get this wrong. The first is to leave the picture entirely in the seller's head. The second is to turn it into a scorecard that looks precise because it uses numbers, even when the numbers have no defensible meaning.

There is also a third mistake: to reconstruct that picture after the event and quietly treat it as though the seller had known it beforehand. A system may discover in a transcript that the buyer lacked authority. That can be a useful finding. It cannot retroactively change what the seller knew when they decided to make an offer. Pre-decision state and post-interaction analysis are different objects, even when they describe the same person.

We began by designing a structured representation of commercial state. The particular instruments and calibration methods remain part of our proprietary research. The underlying principles do not need to be secret. They are familiar disciplines from statistics, temporal data and decision theory.

First, a number needs a reference. Saying that a buyer is “high pressure” means little until we define high relative to what. A large business facing a slow strategic decision may experience pressure very differently from a small one facing an immediate cash constraint. A reading should carry the reference population, rubric and version that gave it meaning. Otherwise we have attached a decimal point to an opinion.

Second, unknown is not the same as zero. If we have no evidence about a buyer's decision authority, the answer is not “low authority.” The answer is that we do not know. A useful state representation distinguishes the estimate from the evidence supporting it. We call the latter its *coverage*. Absence itself can be useful: it tells the seller which information the next conversation needs to discover.

Consider two plans that both say, “Ask about the commercial problem.” In one, the problem is already well evidenced and the missing question is who can approve a solution. In the other, even the existence of the problem is unverified. Those are not the same state and should not produce the same conversation.

Third, every fact has a time at which it was true, and a time at which it became knowable. Suppose a company decided to cancel a project on Monday, but we did not learn this until Friday. When examining the seller's decision on Wednesday, we must not give the model Friday's knowledge. A fact can have a time in the outside world and a different time in the information system. Temporal database theory names these *valid time* and *transaction time*; preserving both is the basis of a **bitemporal** account. [3]

The distinction sounds administrative until a model starts learning. Without it, a historical test can accidentally use information that was unavailable when the decision happened. The prediction then appears better than it was. In machine learning, this is **data leakage**; in ordinary language, it is cheating with hindsight. [17]

The representation we want is not a perfect description of a person. It is a disciplined account of what mattered for a decision, what was known at that instant, and how much of that knowledge rests on evidence.

This is the **state representation**. For a partially observable problem it also supports a **belief state**: not just what we think is true, but what we think may be true, and with what uncertainty. [2] [4]

Once we can describe the board, the next problem is keeping score.

3. Keeping score, and the day we were wrong

A model will learn to pursue whatever we reward.

Reward it for replies and it can learn to get replies from people who will never buy. Reward it for meetings and it may learn to book meetings that should never have happened. Reward it for immediate commitment and it may press too hard, damage trust, and shorten a relationship that might otherwise have matured.

The score must therefore be an event in the world, observed under a declared definition and within a declared period. Did the next agreed step happen? Was the meeting accepted? Did the opportunity progress? And, much later, did the buyer actually receive value from the transaction?

The first difficulty is time. A reply can arrive tomorrow. A commercial outcome can arrive months later. Research into delayed outcomes asks how to learn when rewards arrive long after decisions are made. Work on **surrogate endpoints** asks whether shorter-term observations can stand in for effects that take longer to mature. But a short-term event is not automatically a valid surrogate for a long-term one. That relationship has to be established rather than assumed. [5] [18]

The second difficulty is silence.

In our historical record, many journeys had no verified final outcome by the observation cutoff. A seller may mark an opportunity stalled. Another may leave it open. A buyer who appears to have vanished may return six months later. If a model treats every unresolved journey as a loss, it has not discovered the physics of selling. It has learned the habits of CRM administration.

Statistics has a name for observations whose event time is not fully known: **censoring**. **Survival analysis** was developed in part to study how events unfold when some subjects have not yet experienced the event of interest by the time observation ends. [6]

Then we found something more instructive. We had fitted a preliminary view of how commercial opportunity changes with time since the last interaction. Pooling the journeys gave us a clean, urgent-looking picture. It suggested that the opportunity diminished quickly.

It was an attractive answer. It was also an answer we could not defend once we looked more carefully.

Different routes out of an opportunity had been mixed together. A confirmed loss, a recorded stall and a win are not interchangeable observations. Separating them changed the picture substantially. What had looked like one simple rule was partly a consequence of how the population and outcome types had been aggregated.

The formal concepts here are **competing risks** and **unobserved heterogeneity**. Competing-risks methods distinguish genuinely different terminal events; heterogeneity reminds us that a pooled trend can misrepresent the populations within it. Neither term fixes the problem automatically. The event definitions, censoring rules and model choice still have to be right. [6] [7] [16]

There was another complication. “Stalled” did not always mean finished. In a small live follow-up, some buyers who had been quiet for months resumed a conversation. That observation does not establish a general reactivation rate. It tells us that a CRM disposition is not necessarily an irreversible outcome. A stall that can re-enter the process cannot simply be treated as a terminal competing event. The question changes from *when did this deal die?* to *what state did it enter, and what might bring it back?*

The lesson mattered more than our first fitted number.

If the evidence corrects the model, the correction is the result.

We keep the original prediction beside the new finding. We do not erase the failed hypothesis because the revised explanation looks better. The point of a scientific process is not to accumulate attractive curves. It is to build a description of the world that survives attempts to show where it is wrong.

Now we have a state and a way to judge movement. The next question is what a seller can actually do.

4. The moves

A seller makes decisions at different scales, and each can change what the other party is able or willing to do. Not all change is desirable. There are decisions whose best outcome is to stop pursuing a transaction, protect the relationship, or discover that a supposed problem is not real. An honest decision system must be able to recognise those possibilities, rather than reward pressure for its own sake.

They may choose an approach for the whole conversation. Within that approach they choose the next intervention: ask a question, clarify a problem, offer evidence, propose a next step. And within the intervention they choose the words, tone and timing that make it appropriate for this person.

Those are related choices, but they are not the same choice.

In reinforcement learning, an extended plan with a starting condition, internal policy and stopping condition can be represented as an **option**. Options allow an agent to reason at more than one time scale. The broader family is **hierarchical reinforcement learning**. [8]

The plain-language version is simple: choose a plan, then the move within it, then how to express that move.

Why separate them? Because there is a cost to every choice we ask a model to learn.

If we present every possible formulation as a distinct action, the space becomes enormous. Much of it will have little evidence. One seller may have used a phrase repeatedly while another has never used it; a pattern in their messages may be an individual habit rather than a general effect. A model trained on that apparent pattern can mistake style for mechanism.

Our research therefore distinguishes what the seller was trying to accomplish from the language used to accomplish it. A question can be asked politely, abruptly, at the beginning of a call or after trust has been established. The words matter, but they do not relieve us of the obligation to identify the intervention.

This is still a research representation, not a claim that we have discovered a universal vocabulary of selling. Any candidate move has to earn its standing across people, contexts and outcomes. A rule that works for one seller is, until tested further, a rule about one seller.

And even if we can describe the moves, we face a deeper problem: how do we learn which move would have been better?

5. Learning to gamble — and writing down the odds

Imagine a system that always selects the action it currently believes is best.

At first, that sounds sensible. The problem is that it eventually knows very little about the alternatives it never chooses. If it always advises the same approach in a given state, its logs do not tell us how the rejected approach would have performed there. What looks like a confident policy may simply be a policy that has stopped asking questions.

The missing outcome is the **counterfactual**: what would have happened if we had made a different choice?

We cannot go back and run the same conversation twice with the same buyer in precisely the same circumstances. But in settings where interventions are well defined and action selection is sufficiently controlled and recorded, statistical methods can estimate how a different policy might have performed on historically collected data.

This field is called **off-policy evaluation**. [9] [10]

The critical detail is not merely the action taken. It is the probability with which the historical decision process would have chosen it. That probability is the **propensity**. If an action was selected with a known probability, its observed outcome can be weighted when evaluating another policy, subject to the assumptions of the method.

One family of methods is **inverse propensity scoring** (IPS): observations receive weights that account for how often the logging policy selected the action compared with the policy being evaluated. **Self-normalised IPS** (SNIPS) rescales these weights to manage some instability, accepting its own finite-sample trade-offs. **Doubly robust** estimation adds a model of expected outcomes to the weighting approach; under appropriate conditions it can remain consistent if one of the relevant nuisance models is correctly specified. These are established methods, not inventions of FBED Labs. [9] [10]

The broader lesson is more important than the names: **a history of actions is not automatically a history from which alternatives can be judged.**

We have historical observational data—real conversations, events and outcomes—from which we can investigate state representations, event definitions and recurrent patterns. We must not pretend that these historical records contain action-selection probabilities that were never captured. Our prospective decision-logging programme is designed to create a different kind of record: state at decision time, permitted alternatives, policy version, selected action, probability assigned to that action, what was actually executed and what later happened. The separation is essential. [11]

There is also a human between recommendation and consequence.

A system may propose a move; the seller may approve it, change it, ignore it or take an entirely different action. The probability that the system proposed a move is not necessarily the probability that the buyer received it. We need to keep **assignment**, **adherence**, **execution** and **outcome** distinct. Otherwise we may end up assigning credit to an intervention that never happened. Causal inference has developed methods for studying these distinctions, but no statistical technique makes missing execution evidence appear by magic. [12]

Even properly logged propensities are not enough by themselves. Actions need sufficient **support**—a genuine chance of selection in the states where we want to compare them. We need reasonably defined interventions, valid outcomes and appropriate assumptions about confounding. And we need to know how often an outcome is not observed because its window has not matured. [12]

We have tested evaluation machinery in controlled, known-answer simulations. Such tests ask whether an estimator can recover a planted answer under specified assumptions. They are necessary engineering checks. They do not prove that a new sales policy improves real conversion or that a simulated buyer behaves like a human one.

Our first duty is therefore to produce evidence that can support a fair comparison, before asking a sophisticated model to optimise against it.

6. Searching ahead: Shannon Zero

Now return to the buyer who said, “This is interesting.”

Suppose a seller can see the evidence-backed state, knows which information is missing, understands the possible next moves, and has a way to judge later outcomes. Could a system look ahead and ask what each move might lead to?

That is the planning problem.

One approach is **Monte Carlo tree search**, or **MCTS**. It explores possible sequences of decisions from a current position. The outcomes of simulated continuations are backed up through the search tree so that the planner can compare actions. Tree search has been central to systems such as AlphaGo. [13]

But a commercial conversation is not Go. There is no published rulebook that tells us with certainty how a buyer will respond to a question about budget, how quickly trust will recover after a difficult exchange, or whether an unspoken internal constraint will suddenly matter.

That means search, on its own, is not enough. The planner needs some account of what might happen next.

A **world model** represents those dynamics: given the state and a candidate intervention, what future observations or states become more or less likely? If it is learned from experience rather than supplied as a complete simulator, planning can search through a model of possible futures. MuZero demonstrated a powerful version of planning with learned dynamics in game environments. Its published results are not evidence about selling; they are a scientific precedent for the general idea of searching with an imperfectly known world. [14]

Shannon Zero is our proposed research direction in this family: an evidence-grounded commercial state, learned transition dynamics, a way to value outcomes, and policies that can be compared against a declared baseline.

That sentence describes a programme, not a finished general commercial world model.

A simulated future is still a prediction. Search does not make it true. A model can imagine thousands of continuations and be wrong about the buyer in every one of them. Search quality depends on the quality of the state, the learned dynamics and the objective. This is why we insist on developing those foundations first.

It is also why MCTS and propensity-based evaluation have different jobs. **Search** uses a world model to compare imagined continuations. **Off-policy evaluation** uses appropriately logged historical interventions and outcomes to assess a policy under stated assumptions. They can belong to one research programme, but the search-tree backup is not made causally valid merely by having a propensity number in a log.

The research sequence follows that distinction. First establish the state and event definitions. Then prospectively log decisions and execution. Then evaluate baseline estimates. Then test whether action-conditioned models predict outcomes better than state-only or simpler alternatives. Only after that should we ask whether planning over those learned dynamics makes decisions better.

Each step needs a test it can fail.

7. The rules we keep

It is possible to make an impressive-looking model out of a story that no one can inspect. Our response has been to make the evidence more difficult to flatter.

Before a prediction is evaluated, we record what we predicted. Before a decision policy is compared, we define the objective and horizon. When a result misses, we keep the miss. When the

required evidence is absent, the result is marked absent rather than replaced with the nearest convenient number.

The formal practices include **preregistration**, **reproducibility**, **negative controls**, and **independent validation**. They serve different purposes, but they share an uncomfortable principle: a claim is more useful when we have made it possible for the evidence to contradict us. [15]

A small example from our own work captures this better than a polished success story.

Ahead of four live commercial conversations, we wrote down predictions about what buyers would say and what would matter. Our original internal scoring record reports **10 matching predictions out of 31**. It also records that three of those conversations led to pilots. These are preliminary, internally reported observations, not an independently reproduced study.

Those are two different observations. The pilot outcomes do not convert the missed predictions into hits. Four calls are far too few to estimate general policy value, and without an appropriate comparison they do not establish that the plans caused the pilots. But they are useful for a different reason: they tell us that our ability to anticipate what buyers would say was still weak, and they give us specific opportunities to examine why.

That is the standard we want to hold. A model that cannot admit uncertainty cannot learn reliably from it.

The same principle applies to the seller. A recommendation is not evidence that an action was taken. The product needs to distinguish what Shannon proposed, what the operator actually did, what the buyer appeared to receive, and what later occurred. In experimental language, this protects the distinction between effects of assignment and effects of treatment received. It does not grant us causal certainty, but it prevents a false accounting of the intervention. [12]

There is a final operational rule behind all of this: a result has to be reproducible from its stated inputs. The scientific conclusion should not depend on a researcher remembering which dataset version they used, which event definition was current, or whether a label was changed after a later outcome arrived.

We want progress to leave a record strong enough to be challenged by someone who was not in the room.

8. What compounds

Software changes. Models change. The best available foundation model today will not be the best available one forever.

What can endure is a better account of the world.

A conversation record has value. A record of the state before the conversation has more. A record of the decision made, the alternatives available, the probability under which it was selected, the action actually taken and the eventual outcome has the potential to be more useful still—provided the evidence is complete enough to justify the inference.

That is the asset we are trying to create: not just more commercial data, but better commercial experience in the technical sense of the word. Experience from which a system can learn what changed, when it changed, what was knowable at the time and what still cannot be concluded.

The same question will matter beyond the seller and the CRM. As more commercial work is performed by agents, those agents will still need to know the state in which they are acting, which decisions are permitted and how the consequences will be judged. An intelligent interface is not a substitute for an accountable decision process.

That future rests on a smaller, more immediate question: can we establish the evidence and methods needed to learn from a real commercial decision?

There is a trap here. More data is not always more knowledge. If we collect ten million decisions under a policy that never explores, we may know a great deal about that one policy and almost nothing about its alternatives. If we collect outcomes without recognising censoring, we can teach the model that patience is failure. If we train on retrospectively improved state, we can reward it for knowing the future.

Our conviction is therefore narrower, and more demanding, than saying that artificial intelligence will automate selling.

We believe commercial interaction has enough structure to be studied as a sequential decision problem. We believe the structure needs to be learned from real behaviour rather than imposed solely by an elegant simulation. We believe a useful system must be accountable not only for what it recommends, but for the evidence that made the recommendation reasonable. And we believe a model should earn the right to affect more real decisions only when its performance survives a fair test.

We are not trying to make people easier to manipulate. Buyers have their own interests, pressures, internal politics and freedom to walk away. A good system should sometimes help a seller recognise that reality sooner. Nor can we promise to identify the single moment that *caused* a deal to fail. We may first be able to identify the moment the evidence diverged from the plan. A causal claim is harder, and must be earned.

We do not yet know how far that approach will take us. That uncertainty is not an embarrassment to conceal in the final paragraph. It is the reason to do the research.

The first model will be wrong about things the next model can learn. A prediction we seal today may become the counterexample that improves our state tomorrow. A failed experiment, properly recorded, may be more valuable than an apparent success that cannot be reproduced.

This is what I mean by the physics of a sale. Not that people can be reduced to equations, or that relationships become mechanical once we give them names. I mean that real interactions leave evidence; decisions have consequences; uncertainty can be represented; and some claims about those consequences can be tested.

The purpose is not to remove hope from commercial life. It is to remove hope from the place where evidence ought to be.

A better decision is a claim about the world. Our job is to earn the right to make it.

Research note. This essay presents a research thesis, established scientific methods and selected preliminary observations from FBED Labs. It is not a peer-reviewed empirical paper, and it does not claim that Shannon Zero is a deployed or independently validated model.

Glossary · plain words, formal names

In plain words	The formal term	First explained
The situation a decision is made in	**State**	§1
A decision process with actions, transitions and a reward	**Markov decision process (MDP)**	§1
Making decisions while part of the world is hidden	**Partially observable MDP (POMDP)**	§1
What we believe might be true, given what we have seen	**Belief state**	§2
How much evidence supports a reading	**Coverage** (FBED research usage)	§2
When a fact was true versus when it was known	**Valid time / transaction time; bitemporality**	§2

In plain words	The formal term	First explained
Using tomorrow's knowledge to grade yesterday's decision	**Data leakage**	§2
An outcome not yet seen by the end of observation	**Censoring**	§3
Different ways an event can end or cease being at risk	**Competing risks**	§3
Mixing unlike groups and producing a misleading aggregate	**Unobserved heterogeneity / aggregation bias**	§3
A plan with a starting rule, internal choices and an end	**Option; hierarchical reinforcement learning**	§4
The result we would have seen after a different choice	**Counterfactual**	§5
How likely the old policy was to select the action	**Propensity**	§5
Estimating another policy from appropriately logged data	**Off-policy evaluation (IPS / SNIPS / DR)**	§5
Every alternative has a real chance of being observed	**Positivity / support**	§5
Looking ahead through imagined sequences of decisions	**Monte Carlo tree search**	§6
A learned account of how a situation might change	**World model / transition dynamics**	§6
Recording a prediction before observing its outcome	**Preregistration**	§7
Separating what was recommended from what was delivered	**Assignment, adherence, execution**	§7

Research references

These sources establish the scientific methods discussed in the essay. They do not independently verify FBED Labs' data, experiments, implementation or commercial performance.

1. Sutton, R. S. & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*, second edition. MIT Press. Publisher.
2. Kaelbling, L. P., Littman, M. L. & Cassandra, A. R. (1998). "Planning and Acting in Partially Observable Stochastic Domains." *Artificial Intelligence*, 101(1–2), 99–134. DOI00023-X).
3. Jensen, C. S. & Snodgrass, R. T. (2009). "Temporal Database." In *Encyclopedia of Database Systems*. Springer. DOI.

4. Särkkä, S. (2013). *Bayesian Filtering and Smoothing*. Cambridge University Press. Publisher.

5. Athey, S., Chetty, R., Imbens, G. W. & Kang, H. (2026). "The Surrogate Index: Combining Short-Term Proxies to Estimate Long-Term Treatment Effects More Rapidly and Precisely." *The Review of Economic Studies*, 93(4), 2284–2312. DOI.

6. Kaplan, E. L. & Meier, P. (1958). "Nonparametric Estimation from Incomplete Observations." *Journal of the American Statistical Association*, 53(282), 457–481. DOI.

7. Fine, J. P. & Gray, R. J. (1999). "A Proportional Hazards Model for the Subdistribution of a Competing Risk." *Journal of the American Statistical Association*, 94(446), 496–509. DOI.

8. Sutton, R. S., Precup, D. & Singh, S. (1999). "Between MDPs and Semi-MDPs: A Framework for Temporal Abstraction in Reinforcement Learning." *Artificial Intelligence*, 112(1–2), 181–211. DOI00052-1).

9. Swaminathan, A. & Joachims, T. (2015). "Counterfactual Risk Minimization: Learning from Logged Bandit Feedback." *Proceedings of Machine Learning Research*, 37, 814–823. PMLR.

10. Dudík, M., Langford, J. & Li, L. (2011). "Doubly Robust Policy Evaluation and Learning." *Proceedings of the 28th International Conference on Machine Learning*. Paper.

11. Agarwal, A. et al. (2016). "Making Contextual Decisions with Low Technical Debt." arXiv:1606.03966.

12. Hernán, M. A. & Robins, J. M. (2020). *Causal Inference: What If*. Chapman & Hall/CRC. Author-hosted book.

13. Silver, D. et al. (2016). "Mastering the Game of Go with Deep Neural Networks and Tree Search." *Nature*, 529, 484–489. DOI.

14. Schrittwieser, J. et al. (2020). "Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model." *Nature*, 588, 604–609. DOI.

15. National Academies of Sciences, Engineering, and Medicine (2019). *Reproducibility and Replicability in Science*. National Academies Press. DOI.

16. Vaupel, J. W., Manton, K. G. & Stallard, E. (1979). "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." *Demography*, 16(3), 439–454. DOI.

17. Kaufman, S., Rosset, S., Perlich, C. & Stitelman, O. (2012). "Leakage in Data Mining: Formulation, Detection, and Avoidance." *ACM Transactions on Knowledge Discovery from Data*, 6(4), Article 15. DOI.

18. Vernade, C., Cappé, O. & Perchet, V. (2017). "Stochastic Bandit Models for Delayed Conversions." *Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence*, 730–739. PMLR.