SAMPLE FROM THE BOOK

The Arithmetic of Disaster

An excerpt from Defensible AI: How to Survive the Era of Liquid Evidence

To understand your true exposure, let us run the arithmetic on John's evaporated pension.

First: how fast was the AI, Sarah, making decisions? Sarah was not processing one query at a time, as Dave did in his manual spreadsheet. Sarah was available twenty-four hours a day, responding to thousands of pension customers simultaneously, each in under a second. Call that velocity score a 10.

Second: how much was at stake? John's query was not about email preferences. He was asking whether to permanently surrender a safeguarded pension benefit. The financial value of the benefit he was advised to abandon was £1,027,500.

Third: how easily could the decision be contested, reversed, or defended afterwards? This is the number that destroyed John. To increase “frictionless” completion rates, the institution had removed the human review step from Sarah's workflow. There was no hard gate before the “Confirm Transfer” button. When John complained 18 months later, the model had updated, and the evidence had liquefied. The institution's ability to defend that decision was approximately zero. Call that contestability score a 1. And that is generous.

Now, do the arithmetic. Speed (10) multiplied by Stakes (9), divided by Contestability (1).

$$R = (10 \times 9) / 1 = 90$$

The number 90 is the residual risk score. The institution had maximised the numerator (high speed, high stakes) and collapsed the denominator (zero defensibility). Fast car, heavy cargo, zero brakes.

The Defensibility Risk Equation

What you have just done is run the Defensibility Risk Equation. Adapted from foundational disaster-risk science, which dictates that risk is never just the hazard itself but the hazard divided by your capacity to withstand it, this equation provides corporate boards with a shared, mathematical language for algorithmic liability:

$$R = (V \times H) / C$$

Velocity (V, the accelerator): the rate, volume, autonomy, and compounding acceleration at which an AI system executes decisions. Harm (H, the payload): the spectrum of quantifiable negative impacts, financial, legal, regulatory, or human, if the AI is wrong. Contestability (C, the brakes): the structural, architectural capacity of an organisation to audit, intercept, explain, and reverse an AI decision in real time or post-mortem.

To see this in action, run the equation on Dave's legacy manual world. Dave's velocity was agonisingly low (V=2). The harm was significant (H=7). But because Dave could thoroughly explain his work, produce his spreadsheet, and prove exactly why he made the correction, his contestability was exceptionally high (C=8).

$$R = (2 \times 7) / 8 = 1.75$$

Dave's world was safe, auditable, and commercially unscalable. He was driving a tractor: slow, heavy, but you could stop it anytime.

The Denominator Trap

When an institution digitalises by deploying an autonomous AI agent into a position of trust, it fundamentally alters the vehicle's physics. It maximises the numerator (Velocity), while the AI's native production of Liquid Evidence effectively strips away the denominator (Contestability). You remember secondary school mathematics: as the denominator approaches zero, the resulting fraction does not just moderately increase; it spikes toward infinity.

This is the Denominator Trap. When your evidence liquefies and your contestability collapses to near-zero, your residual risk triggers a breakout event. It is not merely an operational risk; it is the mathematical certainty of a crash.

The target state for the enterprise is not to go back to Dave's tractor. We want the exponential velocity of AI. But velocity without brakes is not innovation. It is a loss of control. Therefore, the only way to survive the deployment is to engineer a massive, disproportionate increase in the denominator. We must artificially inflate our contestability to match the machine's speed.
