



The Cost of Waiting

Precaution, Opportunity and the Ethics of Not Adopting AI

About this paper

This paper examines a decision that most organisations take without ever recording it, which is the decision to wait. Governance writing on artificial intelligence (AI) has concentrated on the harms of deploying a system too early. That focus is reasonable. Much less has been written about the harms of deploying too late, or of not deploying at all, and less still about who carries those harms. The paper is written for boards, executive teams and the advisers who support them, and for policy readers who want to test a precautionary instinct against the evidence available in the early summer of 2023.

It is the sixth paper in the Ecaveo series on the higher-order issues in AI adoption. The series argues that AI adoption is a changing sociotechnical configuration, in which tools, people, incentives and rules shift together, and that its governance must keep pace with it. Governance that watches only the risks of action will fall behind. The configuration moves whether or not a given organisation moves with it.

Editorial balance. Roughly half of the paper reports and weighs published evidence, from the literature on precaution and responsible innovation to the first field and laboratory studies of generative AI at work. About a third is my analysis of what that evidence implies for organisations. The remainder, concentrated in the last two chapters, sets out my own two-sided decision framework, which is proposed for testing and has not been validated.

Legal and regulatory notice. Parts of this paper describe legislation, regulation, regulator activity or court decisions. Nothing in this document constitutes legal advice. The provision of legal advice sits outside the terms of any engagement with the author or with Ecaveo, and the material is presented to support discussion and further review by qualified advisers. Statements about what a source says are attributed to that source. Where I draw a practical implication that a source does not state, the text says so.

HOW TO READ THE EVIDENCE TAGS

Every substantive claim carries a tag stating what kind of evidence supports it. Two tags are combined where a claim rests on both.

PEER-REVIEWED A journal article or peer-reviewed conference paper.

PREPRINT A working paper, preprint or technical report not yet peer reviewed.

STANDARD A published standard or framework, such as ISO or NIST.

POLICY A government, regulator or court document or action.

INDUSTRY A company announcement, press report, analyst or consultancy report.

SCHOLARLY A scholarly book, edited volume or law journal article.

PRACTICE Practitioner observation, including the author's own.

EVIDENCE GAP The available evidence does not answer the question.

ECAVEO The author's own framework or analysis, proposed rather than proven.

CONTENTS

Contents

FRONT MATTER

A note from the author	4
Executive summary	5

PART ONE · THE TWO ERRORS

1	Two ways to be wrong	8
2	The case for precaution	10
3	The limits of precaution	12

PART TWO · BENEFIT AND DISTRIBUTION

4	Responsible innovation as a middle path	15
5	Who gains, and who waits	17

PART THREE · DECIDING

6	A two-sided decision framework	21
7	Implications and open questions	24

END MATTER

References	26
About the author	27

ILLUSTRATIVE MATERIAL

Every organisation, person and scenario described as illustrative in this paper is invented. Published research and official documents are cited and numbered, and invented material is never presented as evidence.

Research closed 27 June 2023. Sources and events are as they stood at that date.

CITATION

Forrest, P. (2023) *The Cost of Waiting: Precaution, Opportunity and the Ethics of Not Adopting AI*. Ecaveo Working Papers, AI Adoption: The Higher-Order Issues, Paper 6. Ecaveo, 29 June 2023. Available at: ecaveo.com.

Comments and corrections are welcome and will be acknowledged in later versions.

PAPERS IN THIS STRAND

- Paper one. *Knowledge With an Expiry Date*. Evidence, Evaluation and Decision-Making When the AI System Keeps Changing. 14 December 2022.
- Paper two. *The Competence Paradox*. When AI Makes the Team Better and the Expert Worse. 21 February 2023.
- Paper three. *The Unit of Adoption*. Why Organisations Adopt Configurations, Not Models. 31 March 2023.
- Paper four. *Everybody's System, Nobody's Risk*. Accountability Gaps in Distributed AI Governance. 28 April 2023.
- Paper five. *Approved Once, Changed Daily*. From Point-in-Time Ethics to Continuous AI Assurance. 31 May 2023.
- **Paper six. *The Cost of Waiting. Precaution, Opportunity and the Ethics of Not Adopting AI*. 29 June 2023.**

A NOTE FROM THE AUTHOR

Counting what did not happen

In advisory work I am often asked to review a decision after it has gone wrong. I am almost never asked to review a decision that was never taken, although some of the most expensive choices I have seen were of that kind. A proposal is parked until the picture is clearer, the minutes record a sensible deferral, and nobody returns a year later to ask what the deferral cost or who paid for it.

Coaching taught me the same lesson from another angle. An athlete held back from a training stimulus for fear of injury is protected from one harm and exposed to another, since fitness that is not built in one block is hard to recover in the next. Good coaches weigh both. They ask what could break, and they also ask what will be lost if nothing changes this month.

I wrote this paper because the conversation about generative AI in the first half of 2023 seemed to me lopsided. The risks of action were described in detail and often vividly, in the press and in board papers alike, and I share many of those concerns. The risks of inaction were mentioned, where they were mentioned at all, as a competitive worry for shareholders. What struck me in the early productivity studies was a different point. If the gains are in fact largest for less experienced staff, then a decision to wait is also a decision about who goes without help, and that makes it an ethical question as well as a commercial one.

I do not argue that organisations should adopt quickly, and some should not adopt at all for some purposes. My claim is narrower. A board that asks only whether a tool is safe enough has asked half of the question, and the other half needs the same evidence, the same rigour and a named owner.

Paul Forrest
Founder, Ecaveo
29 June 2023

TWO PAGES THAT CIRCULATE ON THEIR OWN

Executive summary

Organisations weighing generative AI tend to treat waiting as the safe default, yet not adopting is a decision with costs of its own, and those costs often fall on the people least able to avoid them. Boards should weigh the risk of deploying and the risk of delaying against the same standard of evidence, and set that standard by how easily each choice can be reversed.

The precautionary instinct is sound where harm would be severe and hard to undo, and 2023 has supplied vivid reasons for it, from lawyers sanctioned for filing invented case law to a public warning about extinction risk. The older literature on regulation points the other way. Caution applied to one side of a decision can produce harms on the other, and those harms are easily overlooked because they arrive as benefits that never happened. The first studies of generative AI at work, early and limited as they are, suggest that those benefits may be concentrated among less experienced workers, which makes timing a question of fairness.

1

Waiting is a decision, and it creates risks of its own.

Measures that reduce one hazard can raise others, whether for the same people or for different ones.¹

SCHOLARLY Evidence on the tightening of American drug regulation in 1962 indicated that stricter pre-market testing slowed the arrival of new medicines, a cost that appeared in no safety record.² **PEER-REVIEWED**

3

The earliest productivity evidence points to substantial gains.

One field study reported a 14% average rise in issues resolved per hour.⁵ **PREPRINT** Separate experiments reported faster and better professional writing,⁶ **PREPRINT** and a 55.8% faster completion of a programming task.⁷ **PREPRINT** Each study covers a single setting, none had been through peer review when this paper was written, and the size of the effects elsewhere is unknown. **EVIDENCE GAP**

5

Delay may widen the gap between organisations as well as within them.

One consultancy estimate puts the potential annual value of generative AI in the trillions of dollars.⁸

INDUSTRY Whether early adopters compound their advantage has not yet been measured, and I treat it as a working hypothesis. **EVIDENCE GAP**

2

Precaution remains the right response where harm would be severe and hard to reverse.

Technologies are easiest to change when their effects are least understood, and hardest to change once those effects are known.³ **SCHOLARLY** In June 2023 a New York court sanctioned lawyers who had filed a brief citing cases invented by ChatGPT, a reminder that careless use has real victims.⁴ **POLICY**

4

The gains appear largest for less experienced and lower-skilled workers.

The customer support study reported the greatest gains among novice and low-skilled agents, including a 35% rise in issues resolved per hour for the least skilled, with minimal effect on the most experienced.⁵ **PREPRINT** The writing experiment found that ChatGPT compressed the spread of performance by helping lower-ability participants more.⁶ **PREPRINT** If so, delay costs most those with least say.

6

Responsible innovation and sandboxes offer a middle path, with uneven reach.

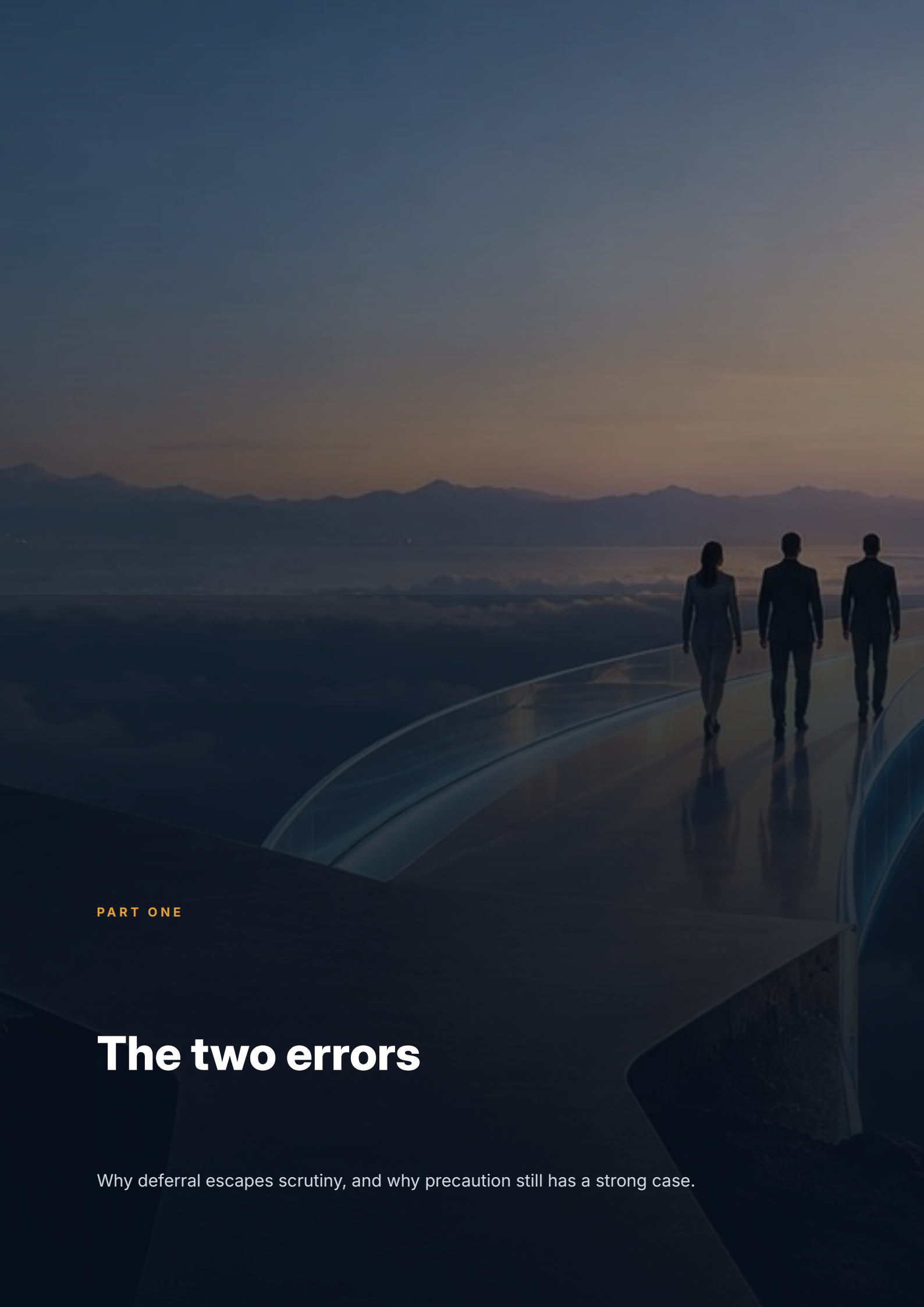
Responsible innovation asks for anticipation and responsiveness throughout development.⁹

PEER-REVIEWED Regulatory sandboxes let firms test products under supervision.¹⁰ **SCHOLARLY** The UK white paper of March 2023 committed to an AI sandbox, but access to such schemes may favour larger organisations.¹¹ **POLICY** **EVIDENCE GAP**

What follows for a board

A board should record a deferral as a decision, with an owner and a review date, together with an estimate of what the deferral costs and who bears that cost. It should ask the same questions of waiting as of acting, and demand strong evidence for any choice that is hard to undo, in whichever direction it runs.

For a use that is minor and easily reversed, a supervised pilot with a dated review is enough. Serious and lasting harm to people calls for independent testing first. In both cases the board should know which staff and customers gain or lose from each option, since the early evidence suggests that those groups may differ.



PART ONE

The two errors

Why deferral escapes scrutiny, and why precaution still has a strong case.

CHAPTER ONE

Two ways to be wrong

An organisation can fail by adopting AI too soon or by adopting it too late, and most governance processes are built to detect only the first failure.

When a board discusses generative AI in 2023, the conversation usually turns on what might go wrong if the organisation deploys it. Will the system invent facts, leak confidential material, embed bias in a decision about a customer or embarrass the firm in public? Each is a fair question, and each has documented instances behind it. The conversation seldom turns with equal force on what might go wrong if the organisation does nothing, even though the two questions concern the same decision viewed from opposite ends.

This paper argues that the imbalance is a failure of governance. Declining to adopt is itself a choice, with consequences that can be estimated and owned, and a precaution that inspects only one side of a decision can leave people worse off than a considered deployment would have done. The argument does not favour speed. It favours symmetry, which means asking of the decision to wait the same questions a board already asks of the decision to act.

1.1 Two firms, both wrong

A pair of invented cases shows the shape of the problem.

ILLUSTRATIVE SCENARIO

Two firms, both wrong

Two mid-sized professional services firms face the same proposal in the spring of 2023. Each could give its junior staff a generative AI assistant for drafting client correspondence and summarising long case files.

The first firm approves the tool within a fortnight. Nobody tests it against the firm's own documents, nobody writes guidance on checking its output, and within weeks a trainee sends a client a summary that cites a regulatory provision which does not exist. The client notices, the partner apologises, and the firm withdraws the tool entirely.

The second firm refers the proposal to a working group, which asks for an assurance that the tool will never produce an error. No such assurance is available, so the proposal is deferred to the following year. Meanwhile its trainees continue to draft by hand, its senior staff spend their evenings correcting that work, and two competitors begin to advertise faster turnaround at lower fees. Nothing happens that any risk register would record.

Both firms made a poor decision, and only the first will ever find out.

The first firm's error is familiar, and the governance literature has a vocabulary for it. The second firm's error is harder to see because its costs take the form of absences, such as hours not saved, checks that stayed manual, trainees who learned more slowly than they might have done, and clients who moved elsewhere without saying why. In my experience the second error is the more common of the two in cautious and well-run organisations, largely because their governance processes were designed around the first. **PRACTICE**

1.2 Why deferral escapes the minutes

Several ordinary features of organisational life push deferral towards being the default. A decision to adopt has a sponsor and a budget line, so if it fails there is someone to question. A decision to defer usually has none of these. It is recorded as a sensible pause, and it is seldom revisited on a fixed date or against stated criteria. Risk registers list hazards that might occur and are poorly suited to recording benefits that failed to occur, while audit functions, for understandable reasons, examine what was done more readily than what was left undone. **PRACTICE**

Personal exposure points the same way. In the boardrooms I have worked with, a director who approved a failed pilot is questioned far more closely than a director who blocked a pilot that would have succeeded, because the second outcome is never observed at all. **PRACTICE** Nobody behaves badly in this arrangement. Each person responds sensibly to the incentives in front of them, and the sum of those sensible responses is a steady bias towards waiting.

What results is an imbalance in accountability, and nothing in it tells a board whether waiting is in fact the safer course. That question can only be answered by setting the two sides next to each other, which the rest of this paper attempts, beginning with the strongest case for caution (and it is a strong case).

CHAPTER TWO

The case for precaution

Precaution earns its place where harm would be severe, hard to reverse and poorly understood, and generative AI in 2023 meets at least some of those conditions.

A fair account of the cost of waiting has to begin with the reasons for waiting, and those reasons are substantial. The precautionary instinct has a long intellectual pedigree, and its best version is a reasoned response to uncertainty about consequences that cannot easily be undone.

2.1 The Collingridge dilemma

David Collingridge's study of the social control of technology still offers the most economical statement of the problem.³ **SCHOLARLY** He argued that a technology is easiest to shape early, when its consequences cannot yet be foreseen, and that by the time those consequences become visible the technology is so embedded in practice, investment and expectation that changing it has become slow and expensive. Early control suffers from a shortage of information. Late control suffers from a shortage of power.

Generative AI fits the pattern closely. An organisation deciding in 2023 whether to adopt a large language model knows little about how the system will behave on its own documents, how staff will come to rely on it, or which suppliers will still be trading in three years. Once workflows and contracts have reorganised around a tool, reversing course costs far more than declining at the outset would have done. The dilemma therefore gives a principled reason to go slowly. I return to this in the framework chapter, because reversibility turns out to be the variable that reconciles precaution with its critics.

2.2 Failure in public view

The first half of 2023 supplied vivid examples of what careless early use can do. In *Mata v. Avianca*, a federal judge in New York sanctioned two lawyers and their firm on 22 June 2023 after they filed a brief that relied on judicial opinions which did not exist and which had been produced with ChatGPT.⁴ **POLICY** The court's opinion and order imposed a penalty of \$5,000 and required the lawyers to write to their client and to each judge falsely named as the author of an invented opinion.⁴ **POLICY** As I read it, the opinion is concerned less with the use of a chatbot than with the failure to check its output and the conduct that followed once doubts were raised, although that is my summary of a court document and readers should rely on the opinion itself.

A different kind of warning came on 30 May 2023, when the Center for AI Safety published a single-sentence statement, signed by many AI researchers and public figures, which held that mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.¹² **INDUSTRY** The statement concerns advanced systems in general. It says nothing about whether a

WHAT THE COURT SANCTIONED

The Mata opinion attaches its sanction to the conduct of the lawyers, including their failure

particular firm should let its staff draft letters with an assistant, yet its effect on board conversations in the weeks that followed was noticeable in my own work, since it lent a respectable vocabulary to a diffuse unease. **PRACTICE**

to verify citations. It was issued five days before the research for this paper closed.

2.3 What precaution gets right

These episodes support two propositions that any fair account should accept. The first is that some failures of generative AI are serious for the people affected and cannot be repaired simply by withdrawing the tool afterwards, since a client misled by a false citation or a customer refused on the strength of a flawed summary has already been harmed. The second is that the information needed to judge such failures in a particular setting is often missing at the moment of decision, which is Collingridge's point exactly.

Where both conditions hold, waiting for better evidence is a defensible choice, and sometimes the only responsible one. The difficulty is that the two conditions do not always hold together. A principle designed for the hard case has a way of being applied to every case, including the many uses of generative AI that are modest in their stakes and simple to reverse.

CHAPTER THREE

The limits of precaution

Caution applied to one side of a decision can create harms on the other, and those harms go unseen because they arrive as benefits that never happened.

A substantial body of work in regulatory economics and risk analysis examines what happens when precaution is treated as costless. Its conclusions predate generative AI by decades, but they transfer with little adjustment.

3.1 Drug lag

Sam Peltzman's evaluation of the 1962 amendments to American drug law is the classic study.² **PEER-REVIEWED** The amendments required that new medicines be shown to be effective before sale, an understandable response to public alarm about unsafe drugs. Peltzman estimated that the stricter regime greatly reduced the flow of new drugs to market, and he found little sign that it had improved the average quality of those that did arrive, so that on his analysis consumers lost more from the medicines they went without than they gained from the added protection.² **PEER-REVIEWED** His estimates have been debated, and I rely here on the structure of his argument more than on its precise magnitudes.

That structure is what matters for AI. A patient who waits for a medicine held up in review appears in no adverse event report, and a regulator judged on visible harms therefore has every reason to lean towards delay. The same logic applies inside a company, where the compliance function is blamed for a failed deployment and seldom for a benefit that was never taken. The phenomenon is usually called drug lag, and its organisational counterpart has no name that I know of.

3.2 Risk against risk

Aaron Wildavsky extended the point into a general account of how societies achieve safety.¹³ **SCHOLARLY** He contrasted anticipation, meaning the effort to prevent harms before they occur, with resilience, meaning the capacity to learn from harms and recover from them as they arise. A society that leans too heavily on anticipation, he argued, gives up the trial-and-error learning from which much of its safety actually comes.

John Graham and Jonathan Wiener gave the idea practical form in their edited volume on risk trade-offs in health and environmental protection.¹ **SCHOLARLY** Its contributors showed, across a series of cases, that an intervention aimed at a target risk can generate countervailing risks, whether by raising the same or a different kind of risk for the same population, or by raising the same or a different kind of risk for another population.

The categories translate readily, and the translation that follows is mine. A ban on generative AI in a customer service team reduces the target risk of an erroneous reply. It may also raise a countervailing risk for the same customers, who wait longer for answers from tired and less experienced staff, and it may shift risk onto junior employees who go on handling difficult cases without support. Whether those countervailing risks outweigh the target risk is an empirical matter that will differ from one team to the next, and a responsible decision has to look for them before it can weigh them. **ECAVEO**

3.3 The paralysis objection

Cass Sunstein's critique is the most direct.¹⁴ **SCHOLARLY** He argued that the strong form of the precautionary principle is incoherent, since risks lie on every side of most regulatory choices. A rule that forbids any step that might cause harm therefore forbids inaction as well, and so it gives the decision-maker no guidance at all. On his account the principle seems to work only because people attend selectively to some risks and overlook others, and he proposed in its place a narrower principle directed at catastrophic harms, combined with a careful comparison of the costs and benefits of each option.¹⁴ **SCHOLARLY**

Precaution cannot serve as a default, because every option, waiting included, carries risk. Sunstein's objection applies with particular force to generative AI, where the risks of use are vivid and widely reported while the costs of non-use are diffuse and rarely counted. A board that invokes caution without stating what caution costs has, on this argument, applied the principle selectively.

None of these authors argues for recklessness, and Sunstein explicitly keeps a place for precaution where catastrophe is possible. Each asks that caution be assessed with the same discipline as action. That is the standard this paper applies, and it raises a practical question the critics answer only in part, namely how an organisation can learn about a technology while still limiting the harm it might do.



PART TWO

Benefit and distribution

Middle paths between restraint and speed, and what the early evidence says about who gains.

CHAPTER FOUR

Responsible innovation as a middle path

Between blanket restraint and unexamined speed sits a family of approaches that let organisations learn while limiting harm, although their reach is uneven.

If precaution and its critics each capture part of the truth, the obvious next step is to look for ways of acting that keep learning going while holding harm within bounds. Two traditions offer such ways, one from the study of science and technology policy and one from financial regulation.

4.1 Anticipation and responsiveness

Jack Stilgoe, Richard Owen and Phil Macnaghten set out a framework for responsible innovation built on four dimensions, namely anticipation, reflexivity, inclusion and responsiveness.⁹ **PEER-REVIEWED** Anticipation asks innovators to consider plausible futures before committing, and reflexivity asks them to examine their own assumptions. Inclusion opens the process to wider voices. Responsiveness asks that the direction of innovation change as knowledge and values change. The framework treats innovation as something to be steered while it develops, which answers Collingridge's dilemma by spreading governance across the life of a technology.

For an organisation the translation is fairly direct, although the authors did not write with generative AI in mind and the application here is mine. Responsiveness implies that an adoption decision should be revisited on a schedule, with named criteria for expanding, pausing or withdrawing. The same logic applies in reverse. A decision to wait should also be anticipated, reflected on and revisited, and the people it affects should have some voice in it, which under present practice they rarely do. **ECAVEO**

4.2 Sandboxes and their reach

Regulatory sandboxes apply a similar logic to regulation itself. Dirk Zetsche and colleagues examined the sandboxes that financial regulators, with the UK's Financial Conduct Authority among the first, created for fintech firms, in which firms test products with real customers under tailored requirements and close supervision.¹⁰ **SCHOLARLY** They regarded the sandbox as one step within a broader move towards what they called smart regulation, and they were cautious about treating it as sufficient on its own.¹⁰ **SCHOLARLY**

The UK government has adopted the idea for AI. Its white paper, presented to Parliament on 29 March 2023, proposes that existing regulators apply five cross-sectoral principles, beginning with safety, security and robustness, and states that the government will not put those principles on a statutory footing initially.¹¹ **POLICY** The white paper gives as part of its reasoning a concern that rigid and onerous legislative requirements could hold back innovation, which is in effect a statement that regulatory

caution has costs.¹¹ **POLICY** It also confirms a commitment to establish a regulatory sandbox for AI, following a recommendation by Sir Patrick Vallance, and its consultation closed on 21 June 2023.¹¹ **POLICY** I describe these proposals as a business adviser reading a policy document, and nothing here is a view on how any future rules will operate.

That limitation bears directly on this paper's concern with distribution. A sandbox tends to favour organisations with the people to apply and to report, and a small firm or a public body with a thin compliance function may find the door open in principle and closed in practice. I have found no evidence that settles the point for AI. **EVIDENCE GAP** If the concern is justified, the middle path shortens delay mostly for those who were least likely to be held back, which is an awkward result for a tool meant to widen access to innovation.

WHO GETS INTO A SANDBOX

Sandboxes select their participants, and applying takes staff time and legal capacity. Whether small firms and public bodies benefit in proportion has not, to my knowledge, been measured for AI.

CHAPTER FIVE

Who gains, and who waits

The earliest evidence suggests that generative AI helps less experienced workers most, which turns the cost of waiting into a question of fairness as well as of competition.

Arguments about the cost of delay are only as good as the evidence on benefit, and until 2023 that evidence was thin. Three studies released in the first half of the year changed the position, although none of them yet settles it.

5.1 The first evidence of benefit

The largest of the three, by Erik Brynjolfsson, Danielle Li and Lindsey Raymond, followed the staggered introduction of a generative AI conversational assistant among 5,179 customer support agents.⁵ **PREPRINT** Access to the tool raised productivity, measured as issues resolved per hour, by 14% on average, and the authors also reported that it improved customer sentiment and employee retention.⁵ **PREPRINT** The staggered roll-out makes the comparison more credible than a simple before-and-after study.

Sida Peng and colleagues ran a controlled experiment in which recruited developers were asked to implement an HTTP server in JavaScript as quickly as possible. Those with access to GitHub Copilot completed the task 55.8% faster than the control group.⁷ **PREPRINT** Shakked Noy and Whitney Zhang, in a preregistered online experiment, gave occupation-specific writing tasks to 444 college-educated professionals and let half of them use ChatGPT; time taken fell by 0.8 standard deviations and the quality of output rose by 0.4 standard deviations.⁶ **PREPRINT**

At the level of the whole economy, consultancy estimates point the same way. McKinsey estimated in June 2023 that generative AI could add the equivalent of \$2.6tn to \$4.4tn a year across the 63 use cases it analysed.⁸ **INDUSTRY** Estimates of this kind rest on assumptions about adoption and task exposure that nobody can yet test, and I treat the figure only as an indication of the stakes.

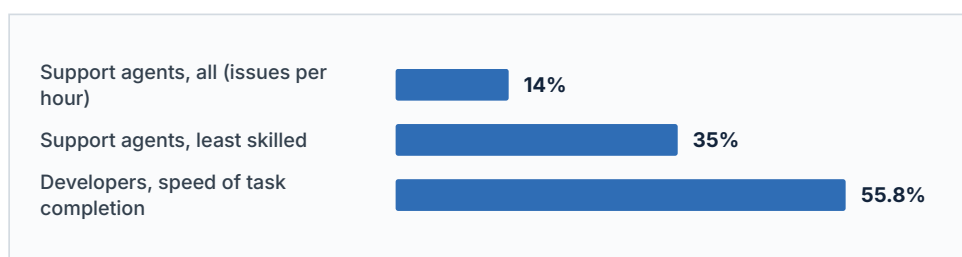


FIGURE 1 Reported gains in two early studies of generative AI at work, on different measures. Sources: Brynjolfsson, Li and Raymond (2023); Peng et al. (2023).

The limits of this evidence need stating early. Each study examined a single task or a single firm, all three were working papers or preprints when this paper was written, and none follows workers over several years. A review of the literature on human and

AI collaboration in organisations, presented at the European Conference on Information Systems in June 2023, frames the creation of value as dependent on organisational conditions.¹⁵ **PEER-REVIEWED** That framing suggests that gains observed in one firm may not transfer to another that lacks similar conditions.

5.2 The distribution of gains

Two of the studies look beyond averages. Brynjolfsson, Li and Raymond reported the greatest gains among novice and low-skilled agents, including a 35% rise in resolutions per hour for the least skilled, and minimal effect on experienced and highly skilled ones.⁵ **PREPRINT** They offered suggestive evidence that the assistant spreads the practices of the most able agents and helps newer agents move down the experience curve. Noy and Zhang found that ChatGPT helped lower-ability participants more, so that inequality between workers fell.⁶ **PREPRINT** Peng and colleagues reported heterogeneous effects that they saw as promising for people moving into software development careers.⁷ **PREPRINT**

Study	Setting and design	Reported pattern of benefit
Brynjolfsson, Li and Raymond (April 2023)	5,179 customer support agents, staggered introduction	Largest gains for novice and low-skilled agents
Noy and Zhang (March 2023)	444 professionals, preregistered online writing experiment	Larger gains for lower-ability participants
Peng et al. (February 2023)	Controlled programming experiment with GitHub Copilot	Heterogeneous effects, promising for career entrants

TABLE 1 Early evidence on who benefits from generative AI. Sources: the working papers and preprint cited in the text.

If these patterns hold beyond their original settings, a decision to wait lands unevenly. An experienced professional loses little from a year's deferral, whereas a trainee or a recently hired agent may lose a good deal, both in daily workload and in the pace at which they learn the job. Within an organisation those are usually the people with least say in the decision, and across an economy they include workers in firms that cannot afford to experiment. On my reading this is the ethical core of the cost of waiting, and it is the part least likely to be written down anywhere. **PRACTICE**

5.3 Stratification

The same logic may operate between organisations. An early adopter that learns to use a tool well accumulates know-how and adapted processes, each of which makes the next step cheaper, while a late adopter starts further back and may find the gap wider than it was when it chose to wait. I know of no study that yet measures this compounding for generative AI, so the claim should be read as my hypothesis.

EVIDENCE GAP

Some considerations run the other way. Late adopters can learn from the mistakes of early ones and avoid sinking money into approaches that fail, which is the hopeful side of Collingridge's argument. Whether learning first beats following probably depends on the sector and on how much of the benefit rests on adaptation specific to

EARLY EVIDENCE

Every productivity figure in this chapter comes from a working paper or preprint released in 2023. The findings point the same way, but their size in other settings is unknown.

the firm (my guess, for what it is worth, is that the second factor matters more than people expect, though I cannot yet show it). A board cannot resolve that question in the abstract, and it may not need to, provided it has some way of noticing when the gap starts to open.

A surreal landscape featuring a city skyline, likely San Francisco, in the background under a hazy, orange-tinted sky. In the foreground, two silhouetted figures in business suits walk along a curved, reflective path that appears to be floating or built on a dark, rocky outcrop. The path has a blueish glow along its edge. The overall mood is contemplative and futuristic.

PART THREE

Deciding

A two-sided framework for adoption decisions and what it asks of boards and regulators.

CHAPTER SIX

A two-sided decision framework

Boards can weigh adoption and delay on equal terms by asking five questions of each and matching the evidence they demand to how easily each choice can be reversed.

The framework in this chapter is my own, assembled from the literature above and from advisory practice. It has not been tested in any systematic way, and I offer it for use and criticism. **ECAVEO**

6.1 Five questions for any adoption decision

Each question is asked twice, once of deploying the tool and once of waiting.

1	Severity	How bad is the worst credible harm, and for whom?
2	Reversibility	How easily can the harm, or the choice itself, be undone?
3	Time-sensitivity	Does the benefit decay, or pass to others, during the wait?
4	Distribution	Who bears each risk, and who receives each benefit?
5	Evidence threshold	What evidence would change the decision, and by when?

FIGURE 2 Five questions, each asked of deploying and of waiting. Source: author's framework.

Severity comes first because Sunstein's narrower principle keeps precaution for catastrophic harms, and a board should know whether it is dealing with one.¹⁴ **SCHOLARLY**

ECAVEO Reversibility follows from Collingridge. A choice that can be unwound cheaply deserves less scrutiny than one that will lock the organisation into a supplier, a workflow or a public commitment, and the same is true of a deferral that forfeits something permanently.³ **SCHOLARLY** **ECAVEO**

Time-sensitivity is the question most often skipped. Some benefits keep, so a firm that adopts a mature tool in two years loses little. Others decay or migrate, as when trainees pass through a development stage without support, or when clients move to competitors and do not come back, and those are the benefits that the reasoning behind drug lag warns about.² **PEER-REVIEWED** **ECAVEO**

Distribution asks the board to name the people on each side. Graham and Wiener's categories of risk transfer are a useful prompt, and the productivity studies give a first indication that the people who benefit from adoption may be different from the people who carry its risks.¹ **SCHOLARLY** **ECAVEO** The final question, on evidence, borrows from responsible innovation. It turns a vague wish for more certainty into a stated condition with a date attached, which makes the decision reviewable.⁹

PEER-REVIEWED **ECAVEO**

6.2 Matching evidence to reversibility

The central proposal is that the evidence a board demands should rise with the difficulty of undoing a choice, and that the rule should apply to delay in the same way as to deployment. A reversible, low-severity use needs only modest evidence and a short pilot. A use that affects people in ways they cannot easily contest, or that places unchecked output before a court or a regulator, needs strong evidence and independent testing first, as *Mata* illustrates.⁴ **POLICY** **ECAVEO**



FIGURE 3 Evidence thresholds matched to reversibility, for deploying and for waiting. Source: author's framework.

The lower half of the grid is where current practice is weakest. A dated pause alongside a small pilot is easy to justify, because it can be ended at any time and it generates the evidence the organisation lacks. An open-ended deferral that lets a cohort of staff pass through their training without support, or lets a market position erode, is a choice that is hard to reverse and should face the same scrutiny as a risky deployment. My own view is that most boards would recognise the distinction at once if it were put to them, and that it is rarely put.

6.3 Owning the cost of delay

Two practices make the framework usable. One is to record every deferral as a decision, with an owner and the criteria that would reopen it, and to estimate the benefit forgone, even roughly, while saying who forgoes it, since an admittedly uncertain figure still disciplines a discussion better than silence. The other is to consult the people most affected on both sides, which in the case of generative AI will often include junior staff whose voices do not usually reach the board.

None of this prevents a board from deciding to wait. It changes what waiting means, from an absence of decision into a decision with an owner, which is the condition under which a later review can learn anything from it. How to estimate a benefit that was never taken remains, I admit, an unsolved measurement problem.

CHAPTER SEVEN

Implications and open questions

The two-sided view asks different things of boards and of regulators, and it leaves several questions that the evidence of mid-2023 cannot answer.

The implications below follow from the framework and are mine. **ECAVEO**

7.1 For boards

A board that accepts the argument of this paper should change its agenda in modest but concrete ways. Adoption proposals should arrive with an assessment of the cost of not proceeding, set out with the same care as the assessment of risk. Deferrals should appear on a register with review dates. Where a proposal is declined, the minutes should say who bears the forgone benefit and whether anything is being done for those people in the meantime.

Boards should also separate safety controls from the decision to deploy. Many of the failures reported in 2023 were failures of verification and supervision, and in *Mata* the court's attention fell on the lawyers' failure to check what they filed.⁴ **POLICY** On my reading, controls of that kind can often be put in place quickly, which makes them a better first response to such failures than a blanket prohibition. **ECAVEO**

7.2 For regulators

Regulators face a version of the same imbalance, since they are held to account for visible harms and rarely for benefits delayed. The UK white paper's concern that rigid rules could hold back innovation recognises this, and its sandbox commitment offers one route to learning under supervision.¹¹ **POLICY** Two further steps would strengthen that approach, on my own view, which goes beyond what the white paper says. Sandbox access could be designed with small firms and public bodies in mind, so that the middle path does not favour the already advantaged, and regulatory impact assessments could state who bears the cost of delay as well as who bears the risk of use. **ECAVEO**

7.3 Conclusion and open questions

The cost of waiting is real, and it is unevenly shared. Precaution remains the right response where harm would be severe and lasting, and the case for it has been made vividly in 2023. The critics of precaution, from Peltzman to Sunstein, show why it cannot be the default, and the first studies of generative AI at work suggest that the people who lose most from delay may be those with the least power over it. A board that weighs both sides against the same evidence, and scales that evidence to reversibility, will still make mistakes, although it will at least be able to see them.

Several questions remain open, and I set them out for further work.

- How should an organisation measure the cost of a benefit it chose not to take, when the counterfactual is never observed?
- Is there ever a duty to adopt a validated tool, and if so, who would owe it and to whom?
- Do regulatory sandboxes reduce the harm of delay for everyone, or mainly for organisations large enough to enter them?
- Will the early finding that less experienced workers gain most survive replication in other tasks, firms and countries?

END MATTER

References

1. Graham, J.D. and Wiener, J.B. (eds) (1995) *Risk vs. Risk: Tradeoffs in Protecting Health and the Environment*. Cambridge, MA: Harvard University Press.
2. Peltzman, S. (1973) An evaluation of consumer protection legislation: the 1962 drug amendments. *Journal of Political Economy* 81(5), 1049-1091.
3. Collingridge, D. (1980) *The Social Control of Technology*. London: Frances Pinter.
4. *Mata v. Avianca, Inc.*, No. 22-cv-1461 (S.D.N.Y. 22 June 2023) (opinion and order on sanctions).
5. Brynjolfsson, E., Li, D. and Raymond, L.R. (2023) *Generative AI at Work*. NBER Working Paper 31161. Cambridge, MA: National Bureau of Economic Research.
6. Noy, S. and Zhang, W. (2023) Experimental evidence on the productivity effects of generative artificial intelligence. Working paper, March 2023. SSRN 4375283.
7. Peng, S., Kalliamvakou, E., Cihon, P. and Demirer, M. (2023) The impact of AI on developer productivity: evidence from GitHub Copilot. arXiv:2302.06590.
8. McKinsey and Company (2023) *The Economic Potential of Generative AI: The Next Productivity Frontier*. 14 June 2023.
9. Stilgoe, J., Owen, R. and Macnaghten, P. (2013) Developing a framework for responsible innovation. *Research Policy* 42(9), 1568-1580.
10. Zetzsche, D.A., Buckley, R.P., Barberis, J.N. and Arner, D.W. (2017) Regulating a revolution: from regulatory sandboxes to smart regulation. *Fordham Journal of Corporate and Financial Law* 23(1), 31-103.
11. Department for Science, Innovation and Technology and Office for Artificial Intelligence (2023) *A Pro-Innovation Approach to AI Regulation*. CP 815. Presented to Parliament, 29 March 2023.
12. Center for AI Safety (2023) Statement on AI Risk. 30 May 2023.
13. Wildavsky, A. (1988) *Searching for Safety*. New Brunswick, NJ: Transaction.
14. Sunstein, C.R. (2005) *Laws of Fear: Beyond the Precautionary Principle*. Cambridge: Cambridge University Press.
15. Raftopoulos, M. and Hamari, J. (2023) Human-AI collaboration in organisations: a literature review on enabling value creation. *ECIS 2023 Research Papers* 381.

END MATTER

About the author

Paul Forrest is the founder of Ecaveo. He advises boards and leadership teams on the adoption and governance of artificial intelligence, drawing on a background in business process reengineering and on hands-on work building machine learning systems, including small, domain-specific language models. He has served as a special adviser to UK government on the Global Entrepreneur Programme and works as a board member and business adviser.

He is also a qualified athletics coach who coaches sprint, hurdle and endurance athletes as a volunteer, and much of his thinking about how people learn, decide and improve comes from the track.

This paper is part of the Ecaveo Working Papers, AI Adoption: The Higher-Order Issues. The views expressed are the author's own.

ecaveo.com

The Cost of Waiting

Ecaveo Working Papers, AI Adoption: The Higher-Order Issues
Paper six of six · 29 June 2023