



The Unit of Adoption

Why Organisations Adopt Configurations, Not Models

About this paper

This paper is the third in a series of six on the higher-order issues in AI adoption, and it is written for boards, executives and the people who run pilots on their behalf. It asks a narrow question with wide consequences. When an organisation says it is adopting a generative AI model, what exactly is it adopting? My answer is that it adopts a configuration, meaning the model together with the interface through which people reach it, the channel and device that carry it, the people using it, the task in hand, the physical setting of the work and the organisational permission that makes relying on it acceptable.

The series as a whole argues that AI adoption is a changing socio-technical configuration whose governance must keep pace with it. This paper supplies the unit of analysis on which that argument depends. A configuration can be described, tested, owned and changed, and once a model leaves the laboratory those are the things a board actually needs to do with it.

Editorial balance. About two-fifths of the paper reports published evidence from information systems research, human factors and the first experimental studies of generative AI at work. About a third is my analysis of what that evidence implies for organisations deciding what to do in the spring of 2023, and the remainder sets out my own configuration model and pilot protocol, which are proposals and are tagged as such throughout.

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PREPRINT A working paper, preprint or technical report not yet peer reviewed.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

1	One model, two sites	8
2	What acceptance models explain	10

PART TWO · TECHNOLOGY MADE IN USE

3	A practice lens on generative AI	13
4	The configuration model	15
5	Modality and embedding	18

PART THREE · GOVERNING CONFIGURATIONS

6	Testing and governing configurations	21
---	--------------------------------------	----

END MATTER

References	24
About the author	25

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 28 March 2023. Sources and events are as they stood at that date.

CITATION

Forrest, P. (2023) *The Unit of Adoption: Why Organisations Adopt Configurations, Not Models*. Ecaveo Working Papers, AI Adoption: The Higher-Order Issues, Paper 3. Ecaveo, 31 March 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

Why I stopped asking which model is best

For most of this year the first question I have heard from leadership teams has been some version of which model to use. It is a fair question. For a few weeks I answered it on its own terms, comparing outputs and pointing to benchmark tables, and the answers were accurate and, as far as I can tell, almost useless, because the teams that took them away went on to have markedly different experiences with the same tool.

The pattern felt familiar from coaching. A session that works well on a dry track in good light can fall apart on a wet road in the dark, with the same athletes and the same plan, and no competent coach concludes from this that the athletes are not ready for training. The coach looks at the conditions, the equipment, the time of day and what else the athlete is carrying that week. I found myself asking the equivalent questions in advisory conversations. Where will people be when they use this? What will their hands be doing? Will their manager think less of them for asking a machine?

Those questions produced better decisions than the model comparisons had, and this paper is my attempt to organise them. I should be clear about its limits. The configuration model is a synthesis of published research and of what I have observed, and I have not tested it as a whole. The scenario is invented. Where the evidence runs out, and on the question of physical setting it runs out quickly, I have said so.

I also suspect that some of what follows will look obvious to anyone who has deployed technology to frontline staff. That is partly the point. The generative AI debate this month is conducted almost entirely in terms of capability, and the older, practical knowledge about how tools reach work seems to have been left behind in the excitement.

Paul Forrest
Founder, Ecaveo
31 March 2023

TWO PAGES THAT CIRCULATE ON THEIR OWN

Executive summary

Organisations evaluating generative AI are comparing models, yet what they adopt is a configuration in which the model is only one of seven elements, so the same model can succeed in one setting and fail in the next. Boards should evaluate, pilot and govern configurations, and should expect results obtained under one configuration to say little about another. The announcements of March 2023, from GPT-4 to Microsoft 365 Copilot and ChatGPT plugins, have been read as moves in a capability race.¹ **INDUSTRY** This paper reads them differently. Drawing on technology acceptance research, practice-based studies of technology in organisations, human factors and the first experiments on generative AI at work, it argues that interface, modality, channel, device, user capability, task, physical setting and organisational legitimacy decide whether a capable model is used at all. The evidence base is young, the most relevant experiments are still preprints, and almost nothing has yet been published on how physical setting or channel changes the value of the same model, so parts of the argument rest on older research and on observation. The practical consequence is a change in what pilots test and what boards ask. A pilot should vary the configuration, record it, and report every result alongside the arrangement that produced it.

1

The unit an organisation adopts is a configuration of seven elements, of which the model is one.

Capability is the only element that model comparisons measure. The other six are usually decided by whoever writes the implementation plan and are rarely recorded. **ECAVEO**

3

The same technology becomes different technologies in use.

Practice research shows groups enacting different technologies from the same software.⁴ **PEER-REVIEWED** General-purpose language models, which have no fixed set of functions, are unusually open to this.

PRACTICE

5

Early productivity evidence is strong, and it is specific to the configurations tested.

Developers with GitHub Copilot finished a programming task 55.8% faster than a control group.⁶

PREPRINT Such studies hold interface, task and setting nearly constant, so their effect sizes should not be carried across to other channels and settings.

EVIDENCE GAP

2

Acceptance models remain useful, but they compress the physical and material circumstances of use into a single perception.

Perceived usefulness and ease of use predict use,² **PEER-REVIEWED** and UTAUT adds social influence and facilitating conditions.³ **PEER-REVIEWED** Facilitating conditions is measured as a belief.³ **PEER-REVIEWED** In my view that is too coarse to tell a pilot team whether a poor result belongs to the model or to the arrangement around it. **ECAVEO**

4

Safe use of opaque AI is achieved by designing the arrangements around it.

Research on sociotechnical envelopment describes how an organisation controlled the inputs, outputs, boundaries and human roles surrounding systems it could not fully explain.⁵ **PEER-REVIEWED** That envelope is organisational work with an owner.

6

The vendors are already competing on configuration.

Microsoft 365 Copilot moves the model into applications people already use, which is a bet on channel and on legitimacy.⁷ **INDUSTRY** In my reading, embedding may also move the decision to adopt away from the organisation and towards the vendor's release schedule. **PRACTICE**

What follows for a board

A board should ask which configuration it is adopting, element by element, where that configuration was tested, and what changed between the pilot and the proposed rollout. It should expect a configuration log for every pilot and an owner for every adopted configuration. Embedded AI features need the same scrutiny as a new tool, because a feature that arrives through a routine software update is still an adoption decision, even when nobody in the organisation has taken it.



PART ONE

The wrong unit

Why evaluating models alone leaves organisations surprised by what happens when AI reaches real work.

CHAPTER ONE

One model, two sites

The same generative model can help one team and frustrate another in the same month, and the difference usually lies outside the model.

In the fortnight before this paper closed, three announcements set the terms of public discussion about generative AI. OpenAI released GPT-4 on 14 March 2023, describing it as a large multimodal model that accepts image and text inputs and produces text.¹

INDUSTRY Two days later Microsoft announced Microsoft 365 Copilot, which places a large language model inside Word, Excel, PowerPoint, Outlook and Teams,⁷ **INDUSTRY** and on 23 March OpenAI introduced plugins that connect ChatGPT to third-party services, opening them initially to a limited group of users and developers.⁸ **INDUSTRY** Most commentary has treated these as moves in a capability race.

Boards are asking their executives a question shaped by the same frame, namely which model the organisation should use and whether it is good enough yet. I think the question is badly posed. What an organisation adopts is a configuration in which the model is one element of seven, and because the other six vary from one part of the business to another, the same model can succeed in one setting and fail in the next with no change to the model at all. **ECAVEO**

1.1 A tale of two teams

The pattern is easiest to see in a single case, and the one below is invented for that purpose.

ILLUSTRATIVE SCENARIO

One model, two sites

A mid-sized distribution business licenses a general-purpose language model for two groups. The customer service team, working at desks in a quiet office, reaches it through a browser tab beside the order system. Staff paste in a customer's email, ask for a draft reply, correct it and send it, and within a few weeks the team lead reports that routine replies take noticeably less time.

The warehouse team is given the same model through the same browser interface, opened on shared handheld scanners with small screens. Pickers wear gloves, the floor is loud, and their questions arise while both hands are full, usually about a damaged pallet or an unfamiliar product code. Typing a question is slow, reading a paragraph on a scanner is slower, and supervisors quietly discourage anyone from standing still to do either. Usage falls away after the first week. Comparing the two sites, the project board concludes that the warehouse staff "are not ready for AI".

The board's conclusion is the one most organisations would reach, and I think it is wrong. Nothing about the model differed between the two sites. What differed was nearly everything else, including the interface (a text box on a small screen), the

device, the physical setting, the task (a reply that can wait a minute against a decision needed now) and the social permission to pause and consult a machine in front of colleagues. Each of these would appear in an implementation plan as a detail, if it appeared at all, and together they decided the outcome. **PRACTICE**

Invented though it is, the scenario will be familiar to anyone who has watched a desk-based tool rolled out to frontline staff. **PRACTICE** What is new in 2023 is the speed at which a single general-purpose model is being offered to every part of an organisation at once, which multiplies the number of configurations a board is adopting without describing.

1.2 The wrong unit of evaluation

If the configuration decides the outcome, then an evaluation that holds the configuration constant and compares models answers a narrower question than it appears to. It may identify the better model for the office. It cannot say whether the warehouse would do better with a different model, with the same model through a headset, or with no model at all.

The same limitation runs through the economic estimates published this month. Eloundou and colleagues estimate that about 80% of the US workforce could have at least 10% of their work tasks affected by large language models, and that about 19% of workers may see at least half of their tasks affected.⁹ **PREPRINT** Economists at Goldman Sachs estimate that generative AI could expose the equivalent of 300 million full-time jobs worldwide to automation.¹⁰ **INDUSTRY**

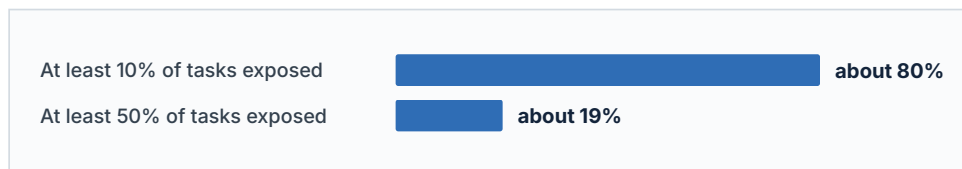


FIGURE 1 Estimated share of the US workforce whose tasks are exposed to large language models, at two thresholds. Source: Eloundou et al. (2023), preprint.

Both figures measure exposure, meaning the share of work that a capable model could in principle touch, and neither measures adoption. Exposure is a property of tasks and models. Adoption happens, or fails to happen, in particular places, through particular devices, among people who have their own reasons to trust or distrust what they are given, and none of that appears in an exposure estimate. **EVIDENCE GAP**

A research agenda for this gap is beginning to form. An opinion paper in the *International Journal of Information Management*, with contributions from more than 70 authors across many disciplines, sets out the opportunities, challenges and research questions that conversational generative AI raises for research, practice and policy.¹¹

PEER-REVIEWED Its breadth is its value, although most contributions still take the tool and its effects as the thing to be studied (my reading, and perhaps an unfair one to some of the authors). What boards need in the meantime is a description of the thing they are adopting that is fine-grained enough to explain the warehouse.

CHAPTER TWO

What acceptance models explain

The dominant models of technology adoption ask whether people find a tool useful and easy to use, and those questions remain necessary for generative AI without being sufficient.

For more than three decades, research on why people adopt information technology has been organised around a small family of models. They are good models. An organisation that ignores them will make avoidable mistakes with generative AI. Their limits matter here for a practical reason, which is that they shape what pilot teams think to measure.

2.1 Usefulness, ease and intention

Davis proposed that two beliefs, perceived usefulness and perceived ease of use, predict whether people will use a system, and he developed measurement scales for both.

² **PEER-REVIEWED** In his studies usefulness was more strongly linked to use than ease of use was, which suggests that people will tolerate some difficulty in a tool that helps them do their job and will drop an easy tool that does not.² **PEER-REVIEWED** For generative AI this is a useful corrective to the attention given to polished chat interfaces.

Venkatesh and colleagues later compared eight competing models and consolidated them into the Unified Theory of Acceptance and Use of Technology (UTAUT).³

PEER-REVIEWED UTAUT places four constructs at its centre, namely performance expectancy, effort expectancy, social influence and facilitating conditions. Gender, age, experience and voluntariness of use moderate their effects. In the authors' validation it accounted for about 70% of the variance in intention to use.³ **PEER-REVIEWED**

Two features of this tradition bear directly on the argument here. Its main outcome is the individual's intention or use, measured largely through that individual's beliefs about the system. Context enters mostly as a moderator, or through facilitating conditions, which UTAUT treats as the degree to which a person believes that organisational and technical infrastructure exists to support use.³ **PEER-REVIEWED** That construct gestures towards the warehouse. It also compresses devices, noise, gloves and supervisory norms into a single perception, which is too coarse to design a pilot around.

2.2 Fit between tool and task

Goodhue and Thompson added a different emphasis. Their technology-to-performance chain holds that a technology improves individual performance when it is used and when it fits the tasks it supports, and their empirical test found that users' evaluations of fit, together with how much people used the system, helped predict the performance effects that users reported.¹² **PEER-REVIEWED**

Task-technology fit moves the analysis closer to this paper's argument, because it treats success as a relation between two things. A generative model that drafts well fits drafting and fits poorly the job of confirming a stock level. The relation has only

two places, though. It says little about the conditions under which the task is done, and the warehouse picker's task in the scenario (deciding what to do about a damaged pallet) is one a language model could in principle help with. The misfit lay in how the capability arrived. **PRACTICE**

2.3 What the models leave out

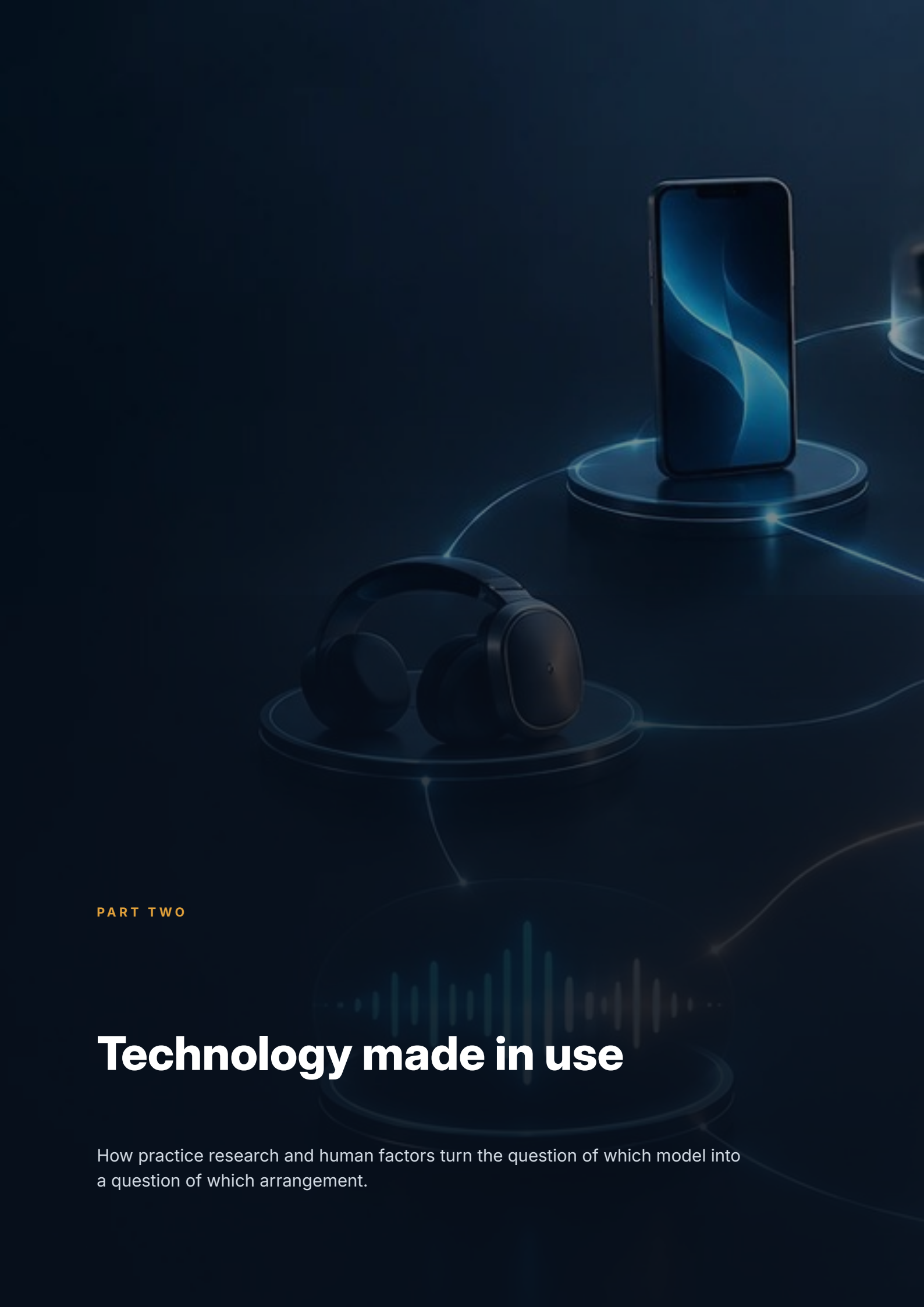
Read together, the acceptance tradition leaves a board with two lessons worth keeping. Perceived usefulness matters more than polish, and social influence and organisational support belong inside the adoption decision from the start.

What the tradition leaves out, at least in the form most practitioners use, is a way of treating the physical and material circumstances of use as design variables. Its surveys ask what users believe about a system they have already met, and they seldom vary the channel, the device or the setting and observe what happens, so they cannot tell a board whether a poor result belongs to the model or to the arrangement around it. **EVIDENCE GAP** Generative AI makes this omission more expensive. The same underlying model can now be reached through a chat window, through an office application, through a plugin inside another service or, in principle, by voice, and each route is a different proposition for the person at the other end.

My own view is that these models are best treated as instruments for measuring one element of the configuration, the user's beliefs, and should not be asked to explain the whole. **ECAVEO** That leaves the rest of the configuration to be explained by other means.

EXTEND OR REPLACE

Acceptance models could in principle add configuration variables as moderators. Whether that would be enough, or whether a different unit of analysis is needed, is one of the open questions this paper ends with.



PART TWO

Technology made in use

How practice research and human factors turn the question of which model into a question of which arrangement.

CHAPTER THREE

A practice lens on generative AI

Practice-based research treats a technology as something that becomes what it is through use, and that view explains why one model can behave like several different technologies.

Information systems research has a second tradition, less often cited in board papers, that starts from what people do with technology. Three contributions from it are directly useful for generative AI, and they build on one another.

3.1 Technologies-in-practice

Orlikowski argued that people enact the structures of a technology as they use it in recurrent practice, so those structures are only partly given by the artefact itself, and she called the resulting patterns technologies-in-practice.⁴ **PEER-REVIEWED** Her field studies of the Lotus Notes groupware found different groups enacting distinct technologies from the same software, depending on their work, their incentives and their understanding of what the tool was for.⁴ **PEER-REVIEWED**

The implication for generative AI is direct. A general-purpose language model is unusually open to this kind of enactment, because it has no fixed menu of functions and responds to whatever it is asked. One team will enact it as a drafting assistant and another as a search engine that sometimes invents its sources, while a third uses it as a patient tutor and a fourth decides it is something to avoid. These are different technologies-in-practice and they carry different risks, even when the licence, the model version and the vendor are identical. **PRACTICE**

3.2 Imbrication of routines and tools

Leonardi takes up a question the practice lens leaves partly open, namely when people change their routines and when they change the technology instead. Studying engineers adopting a new crash-simulation tool at a car maker, he found that people who perceived the technology as a constraint tended to change the technology, while those who perceived new affordances tended to change their routines, and that over time the two kinds of change became interlocked, a process he calls imbrication.¹³

PEER-REVIEWED

Seen this way, the warehouse staff never responded to the constraint at all. Staff could have changed the technology, perhaps by pressing for spoken input, or changed their routines by saving questions for a break. They did neither, because the organisation gave them no authority over the tool and no permission to alter the routine. An organisation that wants imbrication to happen has to leave room for both kinds of adjustment, to the tool and to the routine, and in my experience few pilot plans do.

PRACTICE

3.3 Envelopment

Asatiani and colleagues offer the most practical of the three accounts for a board. Studying how a Danish government agency deployed machine learning systems whose workings could not be fully explained, they describe sociotechnical envelopment, meaning the control of the arrangements around an inscrutable system so that it can be used safely without being fully understood.⁵ **PEER-REVIEWED** In their account the envelope covers such matters as the data the system learns from and receives, the boundaries of what it is allowed to decide, the handling of its outputs and the human roles that surround it.⁵ **PEER-REVIEWED**

Envelopment shows two things. Safe use of opaque AI comes from designing the configuration, since the model itself cannot be made transparent on demand. The envelope is also organisational work that someone must own and revise as the system and its uses change.⁵ **PEER-REVIEWED** Large language models are harder to inspect than the systems in that study and reach far more people, so the envelope has to extend into places the original account had no need to consider, such as the phone in a salesperson's pocket or the scanner in a picker's hand. **PRACTICE**

Taken together, the three accounts converge on a claim the acceptance tradition only brushes against. People adopt a pattern of use, formed by the tool and by the circumstances in which they meet it, and an organisation can shape that pattern deliberately if it knows which circumstances to look at.

A LIMIT OF THE LENS

Practice research is built largely from detailed case studies. It explains variation well and predicts it poorly, which is why this paper turns its insights into variables that a pilot can manipulate.

CHAPTER FOUR

The configuration model

I propose that the unit of adoption be defined as a configuration of seven elements, each of which can be described, varied and governed.

The configuration model is my own synthesis of the research reviewed above and of observation in advisory work. It has not been tested as a whole, and I offer it as a working tool for pilot teams and boards. **ECAVEO**

4.1 Seven elements

The seven elements are as follows.

- Capability, meaning what the underlying model can do and how reliably it does it.
- Interface and modality, meaning how people put requests in and receive results, whether by typed text, speech or image.
- Channel and device, meaning the route and hardware through which the capability arrives, such as a browser tab, an office application, a phone or a handheld scanner.
- User capability, meaning what the people using it know about the task, about the tool and about when to distrust it.
- Task, meaning the work being done, including how quickly an answer is needed and what an error costs.
- Physical setting, meaning noise, light, movement, whether hands are free, and who else can see or hear.
- Organisational legitimacy, meaning whether using the tool is permitted, expected, rewarded or quietly frowned on.

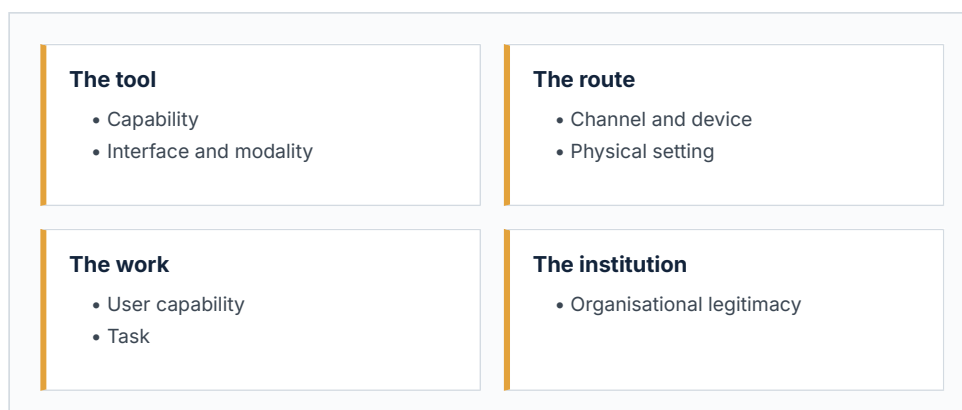


FIGURE 2 The seven elements of an AI configuration, grouped into four clusters. Source: author's framework.

Capability is the only element that model comparisons measure. The other six are usually decided by whoever writes the implementation plan, often implicitly, and they are seldom recorded in a form that would let anyone explain a result later. **PRACTICE** The table sets out what a pilot might ask about each element and what kind of variation would be worth testing.

Element	Question a pilot should answer	Variation worth testing
Capability	Is the output good enough for this task at this cost of error?	Two model versions on the same task
Interface and modality	Can people give and take information in the form the moment allows?	Typed prompt against spoken request and spoken summary
Channel and device	Does the capability arrive where the work already happens?	Separate chat window against an assistant inside the existing application
User capability	Do people know when to rely on the output and when to check it?	Short briefing on failure modes against none
Task	Does the task tolerate delay, and who catches errors?	Drafting against live operational decisions
Physical setting	Can the tool be used here without stopping the work?	Office desk against shop floor or vehicle
Organisational legitimacy	Are people permitted and expected to use it, and seen to be?	Visible manager endorsement against silent availability

TABLE 1 Configuration elements and the questions a pilot should answer about each. Source: author's framework.

4.2 How the elements interact

Most of the surprises in adoption come from interactions between elements. A capable model delivered through a poor channel is experienced as a poor model, and staff will describe it that way in any acceptance survey. Strong legitimacy can carry a mediocre interface for a while, because people persist with a tool they are expected to use, but it seldom survives a setting in which using the tool visibly slows the work. **PRACTICE** User capability can partly compensate for weak capability, since a skilled user who knows a model's failure modes can get value from it that a novice cannot, although that compensation costs attention the task may not have to spare.

Interdependence is why the whole configuration is the unit of adoption. A change to any one element can alter what the other six are worth, so a result obtained under one configuration says little about another. Pilot results should therefore carry their configuration, as laboratory results carry their conditions. **ECAVEO**

4.3 What the first studies show

Early experimental evidence fits this reading, although none of it was designed to test it. Peng and colleagues asked recruited developers to write an HTTP server in JavaScript as quickly as possible, and those given access to GitHub Copilot finished 55.8%

faster than the control group.⁶ **PREPRINT** Copilot is a configuration in the full sense of this paper, since the model arrives inside the programmer's editor, in the form the programmer already works in, at the moment the code is being written.

Noy and Zhang gave 444 college-educated professionals occupation-specific writing tasks and randomly gave half of them access to ChatGPT.¹⁴ **PREPRINT** They report that the tool reduced the time taken and raised the graded quality of the work, and that it shifted effort towards generating ideas and editing and away from rough drafting.¹⁴ **PREPRINT** That last finding is imbrication in miniature.

Both studies are valuable early evidence, and both are preprints. Both also hold the configuration nearly constant, with one task, one interface and a setting chosen by the researchers, which is what good experimental design requires. Their effect sizes therefore describe those configurations. A board should read the 55.8% figure as evidence that a well-configured tool can make a large difference to one well-defined task, and should be wary of carrying it across to work done in other settings through other channels. **EVIDENCE GAP** I know of no published study, as at the date of this paper, that varies the channel or the physical setting of a generative AI tool while holding the model constant.

CHAPTER FIVE

Modality and embedding

Modality and channel look like implementation details, and they often decide whether a capable model is used at all.

Two elements of the configuration model need closer treatment in March 2023, because they are where the practical choices and the vendor strategies are concentrated.

5.1 Why modality matters more than it looks

Human factors research has long held that people cope better with two concurrent tasks when the tasks draw on different mental resources. Wickens's multiple resource theory distinguishes resources by stage of processing, by perceptual modality, by processing code and by visual channel, and predicts that two tasks will interfere more when they make demands on the same resources.¹⁵ **PEER-REVIEWED**

Applied to the scenario, the theory points to a testable hypothesis. A picker scanning shelves is already using eyes and hands, and a text interface asks for more of both. A spoken request with a spoken reply would draw on hearing and speech, which the theory suggests should interfere less with a visual and manual primary task, although a loud floor degrades the auditory channel and the prediction would need testing in exactly that setting. **PRACTICE** (Anyone who has tried to use a voice assistant in a busy kitchen will have a view on this.)

Generative AI has so far reached most workers as text in a chat window, a form that suits people seated at a screen with their hands on a keyboard. OpenAI states that GPT-4 can accept images as input, although image input was not generally available at release.¹ **INDUSTRY** In my view speech recognition and synthesis are now good enough in consumer products to make a spoken configuration plausible for many frontline tasks, but I have not seen that combination evaluated in a workplace study.

EVIDENCE GAP

Modality also changes what counts as a good answer. A paragraph that reads well on a desktop screen is far too long to listen to while working, and a spoken answer cannot easily be skimmed, checked or pasted into a record, so changing the modality changes the work of evaluating the output and with it the error rate the task can tolerate. For a first pilot decision, the distinction between hands-free and hands-busy work will usually tell a team more than the distinction between two models. **ECAVEO**

5.2 Embedding as a configuration bet

On this reading, the March 2023 announcements are less about capability than they seem. Microsoft presents Microsoft 365 Copilot as working alongside people in the applications they already use, and adds Business Chat, which works across the user's calendar, email, chats, documents, meetings and contacts.⁷ **INDUSTRY** That is a bet on channel and device, because the capability moves to where the work already happens.

It is also, on my reading, a bet on organisational legitimacy, since a feature inside a sanctioned application carries an employer's implicit approval in a way that a personal chat account does not. **PRACTICE**

ChatGPT plugins make the complementary move, bringing outside services into the chat interface so that the model can work with current information and act through other applications.⁸ **INDUSTRY** GitHub Copilot, from the Peng study, embeds the model for programmers.⁶ **PREPRINT** My interpretation, which goes beyond anything the vendors say, is that the industry has recognised that capability on its own does not produce use. The vendors are competing on configuration, and above all on channel, while public debate is still conducted almost entirely in terms of the model. **PRACTICE**

5.3 Lower barriers or a hidden choice

Embedding probably lowers the barriers that acceptance models measure, since effort expectancy, which UTAUT defines as the degree of ease associated with using a system,³ **PEER-REVIEWED** should fall when nothing new has to be opened or learned. **PRACTICE** It also has a less comfortable effect.

Embedding a model in an existing tool changes who decides to adopt it.

When the capability arrives as a button in an application staff already use, the choice to adopt moves away from the individual and the pilot team and towards the vendor's release schedule. A board may find that it has adopted a configuration without ever deciding to.

A configuration the organisation did not choose tends to go unrecorded, its envelope has no owner, and when something goes wrong nobody can say which element failed.

PRACTICE Whether embedding lowers adoption barriers or simply hides the choice to adopt cannot be answered on the evidence available in March 2023, and it may well do both, in different proportions for different organisations. **EVIDENCE GAP**



PART THREE

Governing configurations

What pilot teams and boards should do differently once the configuration is treated as the unit of adoption.

CHAPTER SIX

Testing and governing configurations

My recommendations follow from one principle, which is that every pilot, decision and incident report should name the configuration it concerns.

The recommendations in this chapter are my own proposals, derived from the configuration model and from the evidence reviewed earlier, and they have not been tested at scale. **ECAVEO**

6.1 A pilot protocol

1	Describe the baseline	Record all seven elements of the configuration before the pilot starts.
2	Choose one variable	Change a single element at a time, holding the model constant where possible.
3	Match the setting	Run each variant where the work actually happens, at its normal pace.
4	Keep a configuration log	Note devices, settings, model versions and dates, and any change during the pilot.
5	Read results by configuration	Report every outcome alongside the configuration that produced it.
6	Decide and envelope	Adopt a named configuration, assign an owner, and set its boundaries and checks.

FIGURE 3 A pilot protocol for testing configurations. Source: author's framework.

First, test configurations. A pilot that compares two models through one interface in one setting answers a narrow question and should be described that way in any paper that reaches the board. A pilot meant to inform adoption should vary the elements most likely to matter for the work in question, which for frontline work usually means modality, device and physical setting. **ECAVEO**

Second, vary one element at a time where the budget allows. Changing the model, the interface and the device together may produce a better result, and it will leave the organisation unable to say why, which matters as soon as any one of those elements changes again. Where time forces several changes at once, the pilot report should say so plainly. **ECAVEO**

Third, keep a context record. The minimum I would ask for is a description of the seven elements, the model version and the dates of use, and a note of any change during the pilot. Vendors update models and interfaces frequently, in my experience with little notice, and a result that cannot be tied to a version will be hard to interpret a year later. **PRACTICE** **ECAVEO**

Fourth, give each adopted configuration an owner and an envelope, in the sense Asatiani and colleagues describe, covering what goes in, what comes out, what the tool may decide and who checks it.⁵ **PEER-REVIEWED** **ECAVEO** The owner should be the person who would be asked to explain a failure, which is seldom the person who bought the licence.

Fifth, treat embedded features as adoption decisions. When a vendor switches on a generative capability inside an application the organisation already runs, someone should describe the resulting configuration and decide whether to accept it, restrict it or delay it. **ECAVEO**

6.2 What a board should ask

When a proposal to adopt a generative AI tool reaches the board, four questions will usually show whether the configuration has been thought through:

- Which configuration are we adopting, described element by element?
- Where has that configuration been tested, by whom, and in what setting?
- Which elements will differ between the pilot and the proposed rollout?
- Who owns the envelope, and how will they learn when the vendor changes the model or the interface?

A confident answer to the first question and vague answers to the rest usually means the board has been shown a model evaluation and asked to approve an adoption. The two are different decisions. Directors can notice the gap without technical expertise, given the habit of asking where, through what and for whom a result was obtained.

ECAVEO

6.3 Conclusion and open questions

This paper has argued that organisations adopt configurations, and that the same model can succeed or fail depending on the six elements around it. The acceptance tradition explains part of that variation through users' beliefs, practice research explains why the same software becomes different technologies in different hands, and envelopment shows that safe use of opaque systems is a matter of designing and owning the arrangements around them. The first productivity experiments are encouraging and specific to the configurations they tested. The vendor announcements of March 2023 suggest that the industry already competes on configuration, even while the public argument concerns models.

Several questions remain open, and I do not think the evidence available now can answer them:

- Can acceptance models be extended with configuration variables, or does the unit of analysis need to change?
- What is the minimum context record a pilot should keep for its results to be interpretable a year later?
- Does embedding AI in existing tools lower adoption barriers, hide the choice to adopt, or both?
- How far can a board delegate the choice of configuration before it has delegated the adoption decision itself?

The configuration model is a proposal, and I expect some of its seven elements to merge or split once pilot teams start recording them, which is itself a reason to start.

END MATTER

References

1. OpenAI (2023) *GPT-4*. Announcement, 14 March 2023. openai.com.
2. Davis, F.D. (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* 13(3), 319-340.
3. Venkatesh, V., Morris, M.G., Davis, G.B. and Davis, F.D. (2003) User acceptance of information technology: toward a unified view. *MIS Quarterly* 27(3), 425-478.
4. Orlikowski, W.J. (2000) Using technology and constituting structures: a practice lens for studying technology in organizations. *Organization Science* 11(4), 404-428.
5. Asatiani, A., Malo, P., Nagbøl, P.R., Penttinen, E., Rinta-Kahila, T. and Salovaara, A. (2021) Sociotechnical envelopment of artificial intelligence: an approach to organizational deployment of inscrutable artificial intelligence systems. *Journal of the Association for Information Systems* 22(2), 325-352.
6. Peng, S., Kalliamvakou, E., Cihon, P. and Demirer, M. (2023) The impact of AI on developer productivity: evidence from GitHub Copilot. arXiv preprint arXiv:2302.06590, 13 February 2023.
7. Spataro, J. (2023) *Introducing Microsoft 365 Copilot: your copilot for work*. Official Microsoft Blog, 16 March 2023.
8. OpenAI (2023) *ChatGPT plugins*. Announcement, 23 March 2023. openai.com.
9. Eloundou, T., Manning, S., Mishkin, P. and Rock, D. (2023) GPTs are GPTs: an early look at the labor market impact potential of large language models. arXiv preprint arXiv:2303.10130, 17 March 2023.
10. Briggs, J. and Kodnani, D. (2023) *The Potentially Large Effects of Artificial Intelligence on Economic Growth*. Goldman Sachs Global Economics Analyst, 26 March 2023.
11. Dwivedi, Y.K. et al. (2023) "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, article 102642, available online 11 March 2023.
12. Goodhue, D.L. and Thompson, R.L. (1995) Task-technology fit and individual performance. *MIS Quarterly* 19(2), 213-236.
13. Leonardi, P.M. (2011) When flexible routines meet flexible technologies: affordance, constraint, and the imbrication of human and material agencies. *MIS Quarterly* 35(1), 147-167.
14. Noy, S. and Zhang, W. (2023) Experimental evidence on the productivity effects of generative artificial intelligence. SSRN Working Paper 4375283, March 2023.
15. Wickens, C.D. (2002) Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science* 3(2), 159-177.

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 Unit of Adoption

Ecaveo Working Papers, AI Adoption: The Higher-Order Issues
Paper three of six · 31 March 2023