SAMPLE

Building the Autonomous Systems of Tomorrow

Building the Self-Directing Machines of Tomorrow

Aiman JoeHanz

Free sample: Introduction & Chapter 1

putrabooks.github.io

Buy the full book:

Carousell carousell.sg/p/building-the-autonomous-systems-of-tomorrow-ebook-pdf-1465868605

ElevenReader

Ko-fi

Payhip

Contents

Introduction

Chapter 1: The Definition

Chapter 2: The World Model

Chapter 3: The Senses of a Seeing System

Chapter 4: The Planning Mind

Chapter 5: The Motion: Rough Ground

Chapter 6: The Learning Machine

Chapter 7: The Containment: Safety for Systems That Cannot Be Interlocked

Chapter 8: The Crew and the Swarm

Chapter 9: The Human at the Edge

Chapter 10: The Operating Program

About this Series

Connect

This free sample contains the Introduction and Chapter 1. The complete book contains the full contents above.

Introduction

Automation delegates the procedure. Autonomy delegates the judgment. The previous volume built the machine that executes a designed sequence of actions without human intervention at each step: the closed loop, the bounded scope, the exception ladder, the supervised human. This volume builds the machine that goes further — the system released into a world it does not fully control, pursuing a goal through circumstances its designers did not enumerate, deciding for itself what the procedure will be. The self-directing vehicle on the open road, the survey drone over unmapped terrain, the tender robot on the hull, the harvester in the unmarked orchard: systems whose defining act is not execution but decision, made alone, in the field, at speed.

The volume's definition is strict, and its chapters hold to it. An autonomous system is a machine that pursues a stated goal through a world it only partially knows, composing its own actions inside limits its owner has stated, with no human standing inside each decision. Three clauses follow that the whole book builds and defends. A goal, because autonomy without one is drift with a motor. A world that is only partially known, because a fully known world is the automation of the previous volume, where the scope is bounded and the exceptions enumerated. And stated limits, because judgment that is unbounded is not autonomy; it is abdication wearing a machine. The distinction between automation and autonomy is the book's first act — the delegated procedure and the delegated judgment are different delegations, with different failure modes, different supervision, and different debts — and the reader who stops conflating them has already half-built the discipline.

The chapters follow the autonomous system's own anatomy. The world model first — the machine's representation of the world it is moving through, and the uncertainty that must be carried in it, honestly. The senses next — perception rather than measurement: the difference between reading a property and understanding a scene. The planning mind — deliberation and reaction, the hierarchy of plans, and the plan as a first-class artifact. The motion — the rough ground problem, the energy budget, and the stuck state. The learning machine — the components that are trained rather than programmed, and the discipline that keeps them honest as the world drifts.

The containment — the safety question that automation answered with interlocks and that autonomy must answer differently, because judgment cannot be interlocked. The crew and the swarm — the systems that must work together, with and without communication. The human at the edge — trust calibrated, skills kept sharp, the handover that has no good hour. And the operating program — the certification, the simulation-first development, the

field data, and the calendar that keeps a fleet of judgment machines honest for decades.

The volume speaks the same structural language as its predecessors — the definition, the anatomy, the senses, the failure modes, the economics — and adds the one property that makes autonomy harder than every previous subject: the system decides, and the discipline exists so that its decisions remain, provably, the owner's.

Read the chapters in order, from the definition out: Read Chapter 1 first; the definition carries the whole book, and the chapters after it build in order. The reader who operates, buys, or audits an autonomous system can read the chapters as a checklist of what must be true for the system to deserve the name; the reader who designs one will find the same chapters to be the structure of the work, and the reader who merely wants to know what the decade is actually building will find, per the series' standing rule, the parts, the failure modes, and the operating discipline — a builder's account of the self-directing machines, in ten chapters, at the scale of the systems themselves.

Chapter 1: The Definition

The word has been spent. Autonomous has been applied to vacuums and to thermostats, to cruise controls and to chat windows, to every product that does anything on its own — and the spending has left the word nearly worthless as an engineering term. This volume needs it back, and the chapter takes it back the way the automation volume took its word back: with a definition strict enough to audit. An autonomous system is a machine that pursues a stated goal in a partially known world, choosing its own actions within stated limits, without human intervention at each decision. Every clause is load-bearing, and the chapter builds each one before defending the whole.

A stated goal. The autonomous system is going somewhere, doing something, achieving a condition — and the goal is stated, in writing, before the machine is released, because a machine that pursues an unstated goal is not autonomous; it is a hazard with a marketing budget. The goal is to the autonomous system what the procedure was to the automation: the contract's first clause, the thing against which every action is checked. Deliver the cargo, survey the field, return home, hold the position, complete the mission: each is a goal with a completion condition, and the system that cannot say whether it has achieved its goal has not got one.

A partially known world. The automation volume's systems ran against a scope

— the bounded set of conditions the procedure was designed for — and the scope's honesty was that contract's hardest clause. The autonomous system's scope is different in kind: the world is not bounded but modeled, and the model is explicitly incomplete. The road has things on it the map does not; the field has slopes the survey missed; the mission's second half happens in a place nobody has inspected. Partial knowledge is not a deficiency of autonomous systems; it is their defining condition — the reason they must decide rather than execute — and the discipline's first honesty is to state what is known, what is assumed, and what the machine will do when it meets the difference.

Choosing its own actions. This is the clause that separates autonomy from automation, and the volume's opening argument. The automation executes a procedure: its steps enumerated, its branches designed, its response to each exception specified in advance. The autonomous system composes its procedure: it perceives the world, plans a course of action against its goal, executes, perceives again, and replans — and the plan it follows is one it made, within its limits, for the world it actually has. The distinction is not complexity — some automations are intricate and some autonomies are

simple — it is the location of the decision. In automation, the decisions were made at design time by the engineer; in autonomy, some of them are made at run time by the machine. The delegation of execution becomes the delegation of judgment, and every chapter that follows is about what that delegation costs and how it is honored.

Within stated limits. Judgment that is unbounded is not autonomy; it is abdication. The stated limits are the definition's third clause and the discipline's center of gravity: the envelope of the machine's authority — where it may go, what it may do, what it must never do, when it must stop and ask. The limits are stated by the owner, in the same writing that states the goal, and the system that exceeds its limits is not exercising judgment; it is malfunctioning, whatever its intent. The volume's containment chapter builds the machinery that enforces the limits, and the certification chapter builds the evidence that the machinery holds — but the limits themselves are this chapter's contribution: autonomy is judgment inside a fence, and the fence is the designer's, not the machine's.

Without human intervention at each decision. The human remains — the volume's later chapters build that human's whole world — but not inside the loop of every choice. The autonomous system makes its decisions alone, at its own speed, in the seconds between the operator's attention spans, and this is what distinguishes it from the teleoperated machine (a human's hands at a distance) and the supervised automation (a human's approval at each step). The removal of the human from each decision is not the removal of the human from the system; it is the relocation of the human to the ends and the edges, where the volume's supervision chapters put them — and it is the clause that makes autonomy genuinely new, because every previous delegation kept a human inside the cycle.

The automation volume's contract had three clauses: scope, exception paths, and supervision. Autonomy extends each, and the extension is the whole book in three sentences. The scope clause becomes the goal-and-limits clause: the world is not bounded but modeled, and the machine's authority is bounded in the world's place. The exception clause becomes the adaptation clause: the machine does not merely detect that the world has left its scope and stop — it perceives the difference, judges within its limits, and acts, and the ladder of escalation is replaced by a ladder of confidence, with the human summoned when the machine's own uncertainty crosses its stated threshold. The supervision clause becomes the trust clause: the human supervises not the machine's actions but the machine's competence — its record, its calibration, its behavior at the edges of its envelope — because the actions themselves are too many and too fast to supervise one by one.

The extended contract is the book's spine. Chapter by chapter the reader will watch each clause get its machinery: the world model (Chapter 2), the perception (Chapter 3), the planning (Chapter 4), the motion (Chapter 5), the learning (Chapter 6), the containment (Chapter 7), the crew (Chapter 8), the human (Chapter 9), and the program that keeps it all honest (Chapter 10). The contract is not a metaphor; it is the audit. The buyer asks the vendor to show the goal statement, the limits, the model's assumptions, the confidence machinery, and the trust evidence — and the vendor who cannot produce them has not built an autonomous system; they have built a bet.

The discipline's discipline, as in every volume of this series, is the naming of the impostors — the systems that wear the word without the contract. The teleoperated machine: a human's hands at a distance, the surgeon's robot, the bomb disposal arm — remarkable machines, and not autonomous, because the decision never left the human. The automated machine rebranded: the conveyor that sorts, the car that parks itself between painted lines in an empty lot, the drone that flies a stored route — automations, honest and useful, whose delegation is execution and whose vendors have learned that the autonomous price is higher.

The assistive system: the machine that proposes and the human who disposes, the recommendation, the draft, the suggestion — judgment shared, not delegated, and the human who must approve each act is the human who owns it. And the emergent pretender: the swarm, the market, the fleet of simple rules whose collective behavior no one designed — genuinely self-directing, and genuinely outside this volume, because the goal it pursues was never stated by anyone and the limits were never fenced.

Each impostor is a real system with a real discipline, and the naming is not contempt; it is bookkeeping. The teleoperated machine owes its discipline to human factors, the rebranded automation to the previous volume, the assist to the interaction designers, and the emergent to the biologists. What none of them owes — and what this volume is entirely about — is the machinery for delegating judgment to a machine and keeping the delegation honest: the world model, the perception, the planning, the containment, the trust. The reader who can sort any announced system into its true category in one meeting has the definition, and the definition is the chapter.

Between the teleoperated machine and the fully released one lies a ladder, as the automation volume had its rungs of delegation, and the honest buyer asks which rung a system truly stands on. The teleoperation rung: the human decides everything, the machine is the body. The shared-control rung: the human and the machine decide together, the machine proposing and the

human approving, the authority trading hands at designed moments. The supervised-autonomy rung: the machine decides, the human watches, and the machine asks when its own uncertainty crosses the line — the mode of the modern airliner's navigation and the modern farm's harvest. The emergency-override rung: the machine decides everything inside a fenced envelope, the human holds only the stop — the mode of the elevator, the mine, the geofenced drone. And the released rung: the machine decides, no human is in the loop at all, and the envelope is the only supervisor — the mode the decade is negotiating with its regulators.

The ladder's honest lesson is the same as the automation volume's: the rungs are not a maturity scale, and the higher rung is not the better system. Each rung up buys independence and pays in trust: the higher the rung, the more the human must trust the machine's world model, its perception, its planning, and its containment — and the longer and heavier the certification that trust requires. Some missions belong on the bottom rung forever, because the judgment is the value and the machine is only the reach; some belong on the top, because the human's latency is the hazard. The design sin, again, is not the rung but the mislabel: the top-rung claim wrapped around a bottom-rung machine, the human asked to supervise a thousand decisions a second and blamed when they miss one. The ladder is a map of the delegation, and the map must be true.

The chapter closes, per the series' convention, with the audit it argues for. Ask any system that claims autonomy: What is the stated goal, and how does the machine know it has achieved it? What is known about the world, what is assumed, and what happens when assumption meets fact? Which decisions are made at run time by the machine, and which were made at design time by the engineer? What are the stated limits, and what machinery enforces them? And where is the human — inside which loop, at which speed, holding which authorities? A system that answers all five honestly is autonomous in this volume's sense, whatever its domain; a system that answers none of them is a bet wearing the word, and the rest of this book is how the five answers are built, priced, and kept.

The discipline's history is worth one honest section, because the decade's vocabulary was minted by the machines that earned it, and the reader who knows the milestones can hear the engineering inside the words. The species' oldest autonomous artifacts were the mechanical governors and the sea's iron helmsmen: simple, brilliant devices that held a speed or a heading against the world's interference — and the honest lesson they taught was definitional: the governor pursued a goal (a speed) in a partially known world (the load's resistance), without human intervention at each correction. The century that

followed tried to delegate more: the guidance machinery that carried machines across oceans and to other worlds did so with the automation of the previous volume — procedures, timers, sequenced commands — and the deep lesson of those decades was that enumeration works when the world can be predicted, and the world eventually stops cooperating.

The navigation machinery that holds a modern airliner on course across an ocean is the delegation's next rung: the machine chooses its actions continuously within a fenced envelope, the human supervises rather than flies, and the design's safety rests not on the enumeration of states but on the containment of judgment. And the decade this volume describes is the rung above: machines released into the open world — the roads, the fields, the skies — where the world is modeled rather than bounded, the plans are composed rather than stored, and the certification is the negotiation between the fleet's evidence and the public's trust.

The history's arc is the definition's arc: each generation delegated a little more judgment, and each generation had to build a little more machinery to keep the delegation honest — the governor needed only its spring, the ocean needed its procedures, and the open road needs the whole of this book.

The history's cautionary half is equally short and equally structural. The field's most expensive lessons have come from the same error made at each rung:

the delegation claimed a rung higher than the machinery supported. The machine that flew on procedures but was marketed as judgment; the system that decided in the open world but carried the supervision of the workshop floor; the fleet that was released on the strength of its demonstrations rather than the breadth of its evidence. Each failure was, at bottom, a labeling error — Chapter 5 of the automation volume's lesson, arriving at autonomy's altitude — and the discipline's response has always been the same: the ladder's label must match the machinery, the machinery must match the evidence, and the buyer who cannot verify the match is the payer of the gap between them.

The definition's last defense is economic, and the volume states it the way the series always states value: against costs, honestly carried. The delegation of judgment is expensive. It costs the senses that see the open world (Chapter 3), the model that carries it (Chapter 2), the mind that plans in it (Chapter 4), the containment that bounds all three (Chapter 7), the human machinery that supervises the delegation (Chapter 9), and the program that certifies and maintains the whole (Chapter 10) — an estate an order of magnitude heavier than the procedure-execution of the previous volume, and priced accordingly.

The autonomy pays when the mission's conditions make the human's presence impossible or the human's latency ruinous: the dull mission, where the human's vigilance decays and the machine's does not; the dangerous mission, where the human's body is the constraint the mission exists to remove; the distant mission, where the human's logistics cannot follow — the ocean's floor, the storm's interior, the other world's surface; and the dense mission, where the world's speed outpaces the human's loop and the choice is the machine's judgment or no mission at all. The four D's, as the industry's shorthand has it, and the shorthand is honest because it prices the delegation's benefit against its cost.

The economic analysis the buyer should run is the series' standard ledger, extended by one column. The mission's cost with the human in the loop — their presence, their safety, their logistics, their fatigue, their shifts — against the mission's cost with the judgment delegated: the machine's estate, the certification's evidence, the containment's overhead, the supervision's drills, and the residual risk's insurance. The autonomy wins the ledger when the human column's costs dominate — the ocean's floor, the field's poison, the corridor's traffic — and the ledger's honest subtlety is the middle cases: the mission where the human's judgment is occasionally needed, and the design's whole art (Chapter 9's handover machinery) is keeping the human's contribution available at the moments it pays for.

The half-delegation — the machine that handles the routine and summons the human for the rare — is not a lesser autonomy but a designed one, and the fleet that prices its missions honestly will find most of them there: judgment delegated at the margin, authority shared at the edge, and the ledger balanced per mission, not per fashion.

The definition is the book's foundation, and the reader should test it once before leaving. Every chapter that follows is a clause of the definition given machinery: the goal becomes the planner's contract (Chapter 4), the partial world becomes the model and its uncertainties (Chapters 2 and 3), the chosen actions become the mind and the motion (Chapters 4 and 5) and the learned organs that grow them (Chapter 6), the stated limits become the containment

(Chapter 7), the decisions without human intervention become the crew's coordination (Chapter 8) and the human's calibration at the edge (Chapter 9), and the whole delegation's honesty becomes the operating program's calendar (Chapter 10). The reader who holds the definition can read any chapter as the machinery for one of its clauses, and the reader who loses the definition should return here, because every discipline in this book is one of the definition's clauses under load — the audit the definition argues for, and

the audit the decade is learning to demand.

The reader who arrives from the industry's literature will notice what this chapter did not do: it did not grade autonomy on the numbered scale the industry made famous — the levels of driving autonomy, from the assisted to the driverless, minted by a standards committee and quoted by every brochure since. The omission is deliberate, and the section earns it. The levels' honesty is their anchoring: they were written for one domain — the road — and one question — who drives, when — and within that anchor they are a useful shorthand for the ladder this chapter built in its own terms.

The levels' corruption is their travel: applied to every domain, the numbers become a costume — the "level four" farm robot, the "level five" inspection drone — graded by a committee's road-specific definitions onto machines whose roads, drivers, and driving have no referents at all. A harvester has no hands on a wheel to return; a pipeline inspection drone has no driver to send to sleep; the mining loader's teleoperator is simultaneously the "level two" driver and the "level four" passenger, and the number tells the buyer nothing about the delegation's actual shape.

The volume's alternative — the ladder of Chapter 1's earlier section, with its five rungs stated in the definition's own language: who decides, at which timescale, inside which limits, with which summons — is portable across every domain the series covers, because it grades the delegation rather than the machine's domain. A buyer can place any system on it: the teleoperated surgical arm, the geofenced survey drone, the supervised airliner's navigation, the released mine loader — and the placement carries the audit's questions with it, which is the point of any taxonomy in a book about building honest systems. The numbered levels, by contrast, grade the brochure's ambition; the audit's ladder grades the machinery's truth; and the reader who must choose should ask which one a vendor quotes, because the vendor who says "level five" has told the reader about a committee's road, and the vendor who answers the definition's questions has told the reader about their machine.

The volume's last definitional section places the definition in its decade, because the reader deserves the book's honest sense of where the boundary of the discipline currently sits. The technologies of the definition are not the decade's constraint: the senses see, the models model, the planners plan, and the learned organs have crossed the threshold where the seeing and the judging are commercially real — this book's chapters describe machinery that exists and works, per the series' standing rule, with no futures forecast and no vintage denied. The decade's actual constraints are the trust machinery's currencies: the certification's cost (the evidence package's assembly is the

fleet's heaviest engineering artifact, heavier than the machine), the containment's residual risk (priced, insured, and argued over by every regulator whose public shares the machine's world), the commons' tolerance (the manners legislated corridor by corridor), and the human's economics (the operator's estate, drilled and staffed, against the machine's promise of needing no one).

The definition's last clause — without human intervention at each decision — is technologically earned and socially negotiated, and the decade's releases have come where the four currencies balanced first: the mine, the field, the corridor, the warehouse's aisle — the worlds where the risk was priced, the public was absent or consenting, and the evidence was affordable. The reader who wants to know when the definition's machinery comes to their domain should watch the four currencies, not the technology news: autonomy arrives where its trust can be afforded, and the discipline's whole economy is the manufacturing of that trust — which is the operating program's Chapter 10, and the reason this book exists.

And the definition's own last line, held to the convention the volume has kept: a machine that pursues a stated goal in a partially known world, choosing its own actions within stated limits, without human intervention at each decision — read it once as an engineer's specification, and once as an owner's promise, and the volume's ten chapters are the machinery that keeps the two readings the same.

End of free sample. The full book is available at putrabooks.github.io

ABOUT THE AUTHOR

Aiman JoeHanz writes the Building of Tomorrow series, a volume-per-technology account of the systems defining this decade -- each one a structural treatment of the parts, the failure modes, the economics, and the operating discipline a builder can run. In this volume the author takes the series to the delegation itself: autonomy, the machine that composes its own procedure in a world it does not fully control -- treated here as the strict definition and its impostors, the world model and its honest uncertainty, the perception that understands scenes rather than properties, the planning mind and the plan as an examinable artifact, the motion on the rough ground and its budgets, the learned organs and their governance, the containment that bounds what cannot be interlocked, the crew and the commons, the human at the edge and the trust calibrated, and the operating program that keeps the whole delegation honest on a calendar. The catalog applies the same lens to everything else the decade is building.