

The Human Element

Why AI Will Enhance, Not Replace

Sami Karam and Richard Sprague

2025-03-05

Table of contents

Preface	1
Introduction	3
1 The False Binary: Why AI Won't Replace Human Work	9
2 Inside the Black Box: Understanding What AI Actually Does	17
2.1 About AI	17
2.2 How does an LLM-based generative system work?	20
2.3 Statistics of Words	21
2.4 Fine-tuning the output	23
2.5 Wisdom of the crowds	26
2.6 Where do you get the documents	26
2.7 Organizing the data	29
3 The What-How Divide	31
3.1 The Traditional "How" Advantage	31
3.2 AI's Disruption of "How"	32
3.3 The Rise of "What" Skills	32
3.4 Real-World Examples	33
3.4.1 Book-Writing: When Bulldozers Move Words	34

Table of contents

3.5	More examples	36
3.5.1	Software Development: Beyond Code Generation	36
3.5.2	Investment Analysis: Beyond the Numbers	40
3.5.3	AI and Healthcare: Beyond Pattern Recognition	43
3.6	Investment Implications	46
3.7	The Future of Work	47
3.8	Preparing for the Transition	47
3.8.1	The Human Element Remains Central . .	48
3.9	Looking Ahead	48
3.10	Beyond the False Binary	49
3.11	The Nature of the Divide	51
3.12	Philosophical Dimensions of the Divide	53
3.13	Case Studies: The Divide in Practice	54
3.13.1	Software Development	54
3.13.2	Content Creation	55
3.13.3	Healthcare	56
3.14	The Competitive Dynamics of the Divide	57
3.15	Organizational Implications	58
3.16	Investment Implications	60
3.17	The Evolution of Knowledge Work	61
3.18	Integration with Enhancement Thesis	62
3.19	Conclusion: Navigating the Transformation . . .	63
4	Beyond Computation: The Philosophy of Human Intelligence	65
5	The Human Edge	71
5.1	The Uniqueness of Human Judgment	71
5.2	The Architecture of Judgment	73
5.3	The “What-How” Divide in Professional Contexts	75
5.4	Decision-Making Under Uncertainty	76
5.5	The Enhancement Framework Revisited	78

Table of contents

5.6	The Philosophical Stakes	79
5.7	Investment Implications	81
5.8	Judgment as Competitive Advantage	82
6	Finding the Sweet Spot	85
6.1	The Enhancement Zone	86
6.2	The Enhancement Framework	87
6.2.1	1. The “What” versus “How” Distinction .	87
6.2.2	2. The Decisioning Framework and Four- Quadrant Enhancement Model	89
6.3	Real-World Enhancement Sweet Spots	91
6.3.1	Healthcare: Augmented Diagnostics . . .	91
6.3.2	Financial Services: Enhanced Risk Assessment	92
6.3.3	Manufacturing and Transportation: En- hanced Human Capabilities	93
6.3.4	Creative Industries: Collaborative Design	94
6.4	Implementation Principles for Finding Your Sweet Spot	95
6.4.1	The Role of Management and Cultural Considerations	95
6.4.2	Start with Human Needs, Not AI Capabil- ities	96
6.4.3	Design for Appropriate Division of Labor	96
6.4.4	Build Trust Through Appropriate Trans- parency	97
6.4.5	Evolve Through Iteration	98
6.5	Finding Your Organization’s Sweet Spot	98
6.5.1	1. Process Inventory and Mapping	99
6.5.2	2. Pilot Selection and Design	99
6.5.3	3. Capability Building	100
6.5.4	4. Scaling and Evolution	100

Table of contents

6.6	Beyond Optimization: The Strategic Implications of Enhancement	101
6.6.1	Talent Attraction and Retention	101
6.6.2	Sustainable Competitive Advantage	102
6.6.3	System Resilience	102
6.7	Looking Forward: The Human-Centered Future of AI	103
7	The Implementation Challenge: Making Enhancement Work	105
7.1	The Enhancement Framework	106
7.2	Building Trust Through Transparency	107
7.3	The Training Challenge	107
7.4	Measuring Success	108
7.5	Common Implementation Pitfalls	108
7.6	The Path to Successful Implementation	109
7.7	Looking Ahead: The Future of Implementation	113
7.8	The Human Element in Implementation	113
7.9	Investment Implications	114
7.10	Conclusion	114
8	The Human Element in Creative Work	117
8.1	The Beethoven Challenge	117
8.2	The Results: Technical Success, Artistic Failure	118
8.3	The Role of Human Struggle	119
8.4	Beyond Music: The Broader Implications	120
8.4.1	Sports and Entertainment	120
8.4.2	Business Leadership	120
8.4.3	Professional Services	120
8.5	The Enhancement Opportunity	121
8.6	Looking Forward	121

9	Following the Money	123
9.1	The Enhancement Premium	123
9.1.1	Higher Productivity Per Employee	124
9.1.2	Improved Capital Efficiency	125
9.1.3	Greater Customer Retention	125
9.1.4	More Sustainable Competitive Advantages	126
9.1.5	Lower Regulatory Risk	127
9.2	Value Creation vs Value Capture	128
9.2.1	1. Infrastructure Providers	128
9.2.2	2. Enhancement Enablers	129
9.2.3	3. Enhanced Incumbents	130
9.3	Implementation Risk	130
9.4	Valuation Considerations	131
9.5	Timing Considerations	132
9.6	Geographic Considerations	132
9.7	Regulatory Environment	133
9.8	Investment Strategy Implications	134
9.8.1	Sustainable Competitive Advantages vs. Technical Leadership	134
9.8.2	Clear Enhancement Use Cases	135
9.8.3	Data Assets and Implementation Capabil- ities	136
9.8.4	Timing and Geographic Diversification .	137
9.8.5	Regulatory Development Monitoring . .	137
9.8.6	Implementation Framework	138
9.9	Conclusion	138
10	Building the Future: A Human-Centric Vision for AI	141
10.1	The Enhancement Imperative	141
10.2	Rethinking AI Implementation	142
10.3	Policy Imperatives	143
10.3.1	1. Preserving Human Agency	144
10.3.2	2. Promoting Transparency	144

Table of contents

10.3.3 3. Protecting Privacy and Data Rights . .	144
10.4 Investment Implications	145
10.5 The Path Forward	146
10.5.1 For Business Leaders:	146
10.5.2 For Policymakers:	146
10.5.3 For Society:	147
10.6 Conclusion	147
Summary	149
Building the Future: A Human-Centric Vision for AI	151
Author Dialog	154
11 About the Authors	159
References	161

Preface

This is an initial placeholder for a 75,000 word book.

Current word count = 25839

This is the preface

Introduction

In 2021, a team of music historians, musicologists, composers and computer scientists gathered in... to undertake a task that had never been attempted before: to complete Beethoven's tenth symphony with the help of Artificial Intelligence. Beethoven had started composing the symphony but had died in 1827 before making much progress. He only left behind a few musical sketches, not a substantive draft on which the team could easily build. Nonetheless, the team trained a computer on Beethoven's entire body of work and then let AI do the rest of the work, the composing of the entire symphony.

This is the typical way that AI works, in two main steps: first "to learn" and then "to do". An AI program has to first learn or "train" from an existing database everything that it can learn that is relevant to the planned project. This is the training phase. Then the AI program has to do or respond to a request to use what it has learned during the training phase to deliver a response or an output. This is the inference phase.

This is exactly what the ... team did. They trained AI on Beethoven's symphonies and other musical pieces, as well as the early sketches he had started for the tenth symphony, and then asked the AI program to create the rest of the symphony based on what it had learned during the training phase.

We start with this project because it neatly summarizes common aspirations about AI. Many people do expect AI to eventually re-

Introduction

place the vast majority of human work. And here was an example with a team of professionals training a machine to replace, or at least replicate, the work of a human, albeit a genius and not an ordinary human.

There was much pomp and expectation around the project and the team announced that the tenth symphony created by AI would be released in a world premiere performance in Bonn, Germany, on October 9, 2021. When the day came, ...

In the subsequent days and months, a consensus gradually formed about the AI generated tenth symphony. And it was this: the music did sound like Beethoven in many parts, but it lacked the ultimate ingredients that made other symphonies masterpieces: passion, spirit, a tangible human touch. Instead the AI tenth sounded mechanical and betrayed its genesis by being too repetitive in the wrong places. In short, it was precise and competent in the way of a robot, but it was ultimately deficient for its inability to convey passion, to elevate and to inspire listeners.

In our readings about the AI generated symphony, we looked at a large number of opinions by experts from all walks of life. But one in particular from Jan... summarizes, in our view, the central reason that the AI-generated tenth symphony fell short of actual Beethoven-created symphonies. It was not only that it was mechanical and repetitive. Perhaps these flaws could have been fixed in a subsequent iteration. Nor was it that... In the end, as J... put it, the tenth symphony fails to stir us for one main reason which is that listeners “want to see the human do it.”

Humans want to see other humans do it. These eight words sum up the thesis of this book. And from them, we derive our belief that AI can be a very effective assistant to humans in a large

number of tasks but it will never satisfactorily replace humans. Leadership and teamwork require the input and inspiration imparted by a leader and by team mates, and a robot, or AI-trained program, will never be able to provide that human input and this inspiration.

This is the right time to have this conversation because the AI revolution has spawned two competing narratives, both fundamentally wrong. The doomsayers warn of widespread job displacement as artificial intelligence becomes increasingly capable of performing human tasks. The techno-utopians promise a future where AI solves humanity's greatest challenges, freeing us from mundane work. But the reality emerging from actual AI implementations tells a different story – one where artificial intelligence enhances rather than replaces human capabilities. This means that the doomsayers are wrong because the optimal completion of a task will remain elusive without the work of humans using AI. And it also means that the techno-utopians are also wrong because AI will not harness the creativity that is needed to solve on its own our greatest challenges.

Our own view is that AI is best seen as a force multiplier, and a formidable one. Humanity has developed many force multipliers through its history and they have all added to the productivity of humans and to the wealth of society. And now humanity will develop yet another, and perhaps the most powerful of all, force multiplier that will similarly boost productivity and raise living standards.

This is not to say that all jobs will be preserved. Of course, there will be much disruption and dislocation. Many jobs will completely disappear or will be completely taken over by AI, but AI will be in most cases a very competent and very efficient assistant, not a leader or conductor of the proceedings. It is true at the same time that large numbers of people will need to retrain

Introduction

to do other jobs, but this has often be true in our modern society. It may well be that the numbers will be larger this time around, but they were also larger last time than they had been the previous time. Human progress is inherently destructive of jobs that can be replaced by machine. There are millions of robots today that perform the tasks that used to be done by human factory workers only a few decades ago. This has always been the march of technology. AI may be faster than previous revolutions, but then its rewards will also be faster and richer.

Talking more practically, we have already observed a pattern across industries that supports our thesis: the most successful AI applications are those that augment human judgment rather than attempt to replicate it. Here we draw on our long experiences in finance and technology. We see that from financial trading desks to hospital diagnostic centers, from military command posts to creative studios, the winning formula consistently involves keeping humans “in the loop” while leveraging AI’s computational capabilities. We are in the early days of AI but our view is that this will remain true with future AI systems that are far more developed than today’s.

Previous technological revolutions followed similar paths. There was much concern among bank tellers when ATMs were first introduced that the job of a human teller was going to disappear. This was a rational fear. Why wait in line to see a teller if you can just make deposits and withdrawals at an ATM? And from the bank’s point of view, why employ tellers when you can deploy ATMs at a much lower cost? It turned however that people still needed human tellers for transactions that went beyond simple deposits and withdrawals. But another dynamic proved even more revealing. Because banks were able to reduce the number of tellers per branch and were also able to reduce their costs thanks to other technological advances,

they were ultimately able to open more branches, each of which required a minimal number of human tellers. The net result is that the total number of people employed as tellers was in fact larger after ATMs were introduced than before they were introduced.

The same experience was repeated in other fields. Computer-aided design tools were expected to replace architects; instead, they enhanced architects' ability to explore creative possibilities through new technology and it allowed architecture firms to hire more architects.

One key difference today may be the pace and scope of change that AI enables.

What makes AI unique is its ability to process vast amounts of data and recognize patterns that humans might miss. But this capability, impressive as it is, remains fundamentally different from human intelligence. AI can analyze millions of medical images to flag potential anomalies, but it takes a doctor's judgment to interpret these findings in the context of a patient's overall health. AI can process thousands of financial data points per second, but it takes a human analyst to understand how changing geopolitical dynamics might affect market sentiment.

This book challenges both the fear-mongering and the hype around AI. Instead, we present a framework for understanding how AI can enhance human capabilities across industries. Drawing on real-world case studies and our own experience implementing AI solutions, we demonstrate why keeping humans central to decision-making leads to better outcomes than pursuing full automation. For business leaders, this book offers practical guidance on implementing AI in ways that augment rather than replace human workers. For investors, we provide frameworks for evaluating AI companies based on

Introduction

their approach to human-AI collaboration. For policymakers, we outline principles for governing AI development while preserving human agency and judgment.

The coming decades will see artificial intelligence transform every industry. But this transformation will not follow the simple pattern of automation and replacement that many predict. Instead, we are entering an era of enhancement, where human capabilities are amplified by AI rather than superseded by it. Understanding this distinction – and its implications for business strategy, investment decisions, and policy choices – will be crucial for navigating the AI revolution.

The future belongs not to those who try to replicate human intelligence, but to those who find ways to enhance it. This is the human element in the AI revolution.

1 The False Binary: Why AI Won't Replace Human Work

How both AI doomsayers and utopians miss the fundamental nature of human-AI interaction

When ATMs rolled out across bank lobbies in the 1970s, the forecast was grim for tellers. Machines that could spit out cash and swallow deposits seemed poised to erase a whole category of jobs. Why pay humans to handle transactions when steel boxes could do it faster and cheaper? Yet, the numbers told a different story. By the 1990s, the U.S. had *more* bank tellers than before ATMs arrived—not fewer. The machines slashed the cost of running a branch, so banks opened more of them. Tellers didn't vanish; their work morphed into something less mechanical—advising customers, solving problems, building trust. Technology didn't replace them. It redefined them.

This story isn't an outlier. It's a blueprint. From looms to assembly lines, every leap in automation has sparked the same debate: will machines liberate us or leave us jobless? Today, artificial intelligence has reignited that question with a vengeance. Techno-optimists paint a future where AI cures diseases and halts climate disasters, while doomsayers see a world of shuttered offices and idle hands. Both sides, though, are stuck in a

1 The False Binary: Why AI Won't Replace Human Work

false binary—replacement or utopia—that misses what's really unfolding. AI isn't here to take over human work. It's here to amplify it. The evidence, from history to the latest deployments, shows that the most powerful outcomes come when humans and machines collaborate, not when one tries to oust the other.

This chapter sets the stage for that argument. We'll unpack why the replacement narrative keeps falling short, explore where AI shines and where it stumbles, and show how this shift toward enhancement is already reshaping industries. The goal isn't to dismiss AI's potential or its disruptions—it's to reframe the conversation around what it can realistically do alongside us.

Beyond Replacement: The Historical Lens

The ATM tale is a good starting point because it's concrete. Between 1970 and 2010, teller jobs grew from about 500,000 to 600,000 in the U.S., even as ATMs ballooned from a handful to over 400,000. Banks didn't ditch humans; they leaned on machines to handle the rote stuff—counting bills, processing checks—freeing tellers to tackle thornier tasks like mortgage advice or fraud disputes. The tech didn't erase the human touch; it made it more valuable by stripping away the mundane.

This pattern echoes across decades. When computer-aided design (CAD) software hit architecture firms in the 1980s, skeptics predicted drafting tables would gather dust and architects would fade into obsolescence. Instead, CAD turbocharged creativity. Architects could test wilder ideas, tweak designs in real time, and pitch clients with vivid 3D models. Firms didn't shrink—they expanded, hiring more talent to dream bigger. The software didn't supplant human vision; it gave it wings.

Fast-forward to the 21st century, and the script holds. Amazon's warehouses buzz with robots zipping packages along conveyor

belts, yet human workers haven't vanished. In 2023, the company employed over 1.5 million people—up from 800,000 in 2019—despite pouring billions into automation. Robots excel at fetching boxes, but humans handle the exceptions: a torn label, an odd-shaped item, a last-minute order tweak. The machines crank through the predictable; people wrestle with the messy. Together, they've turned Amazon into a logistics juggernaut—not by replacing workers, but by retooling their roles.

These examples cut through the hype. Technology doesn't follow a straight line from human to machine. It zigzags, finding new ways to mesh with what we're good at. AI, for all its dazzle, fits this mold—not breaks it.

The Hype and the Hope: AI's Modern Moment

Enter ChatGPT in late 2022. Overnight, AI went from a buzzword to a living room guest. It could write poems, debug code, even fake a job interview. CEOs scribbled AI-first strategies, stocks like NVIDIA spiked, and headlines swung between marvel and panic. Some saw a golden age dawning—AI as the ultimate problem-solver. Others braced for collapse—white-collar jobs vaporized by algorithms. The truth, as usual, is less dramatic but more interesting.

Take Microsoft's CoPilot, an AI baked into Office tools. A Fortune 500 consumer goods company rolled it out in 2023, hoping to slash grunt work. It churned out email drafts, meeting summaries, and slide decks at lightning speed. But the shine faded fast. Employees spent hours tweaking outputs—fixing tone, adding context, catching errors the AI couldn't see. A manager drafting a client pitch found CoPilot's version polished but flat, missing the rapport that seals deals. The tool saved time on mechanics, sure, but the human layer—judgment, intent, finesse—still ruled the outcome.

1 *The False Binary: Why AI Won't Replace Human Work*

This isn't a knock on CoPilot. It's a clue. AI's strength lies in crunching what's known—data, patterns, templates. It's a wizard at “how” once you've nailed the “what.” But deciding *what* matters—strategy, purpose, nuance—that's where humans hold court. In software development, GitHub Copilot spits out code snippets with eerie precision, but it's useless without a programmer framing the problem: What's the user need? How should this system scale? The AI executes; humans steer.

Even in creative turf, the story tracks. Recall the 2021 bid to finish Beethoven's 10th symphony with AI. A team fed it every note Beethoven ever wrote, plus his early sketches, and let it compose. The result debuted in Bonn to fanfare—and fell flat. Critics like Jan Swafford pegged it right: it mimicked Beethoven's style but lacked his fire. The notes aligned, but the soul didn't stir. Listeners craved the human behind the music—not just the sound, but the struggle and spark that made it real. AI could assemble; it couldn't inspire.

Where AI Shines, Where It Stalls

To get why this keeps happening, we need to peek under AI's hood—just enough to see its edges. Today's systems, like the large language models powering ChatGPT, are pattern machines. They're trained on mountains of text, images, or whatever you feed them, then predict what comes next based on stats and probabilities. Chess-playing AI doesn't “think” like a grandmaster—it calculates moves at a scale humans can't touch. Medical imaging AI doesn't “diagnose”—it flags oddities in X-rays faster than a radiologist's eye.

This is potent stuff. In 2023, a Stanford study found AI could spot pneumonia in chest scans with 92% accuracy, edging out human averages. Financial firms now use algorithms to sift through earnings calls and news feeds, catching signals

in seconds that once took analysts days. But here's the catch: these wins are narrow. The imaging AI can't ask a patient about symptoms or weigh their stress levels. The trading bot can't sense a CEO's bluff or predict a geopolitical curveball. They're tools—sharp, fast, tireless—but not minds.

Humans bring what AI can't: context, curiosity, agency. A doctor doesn't just read scans; she reads people—piecing together lifestyle, history, and gut hunches. A trader doesn't just crunch numbers; he reads the room—gauging fear, greed, or a rival's next move. AI lacks that “being-in-the-world” quality Martin Heidegger flagged in philosophy—our knack for living through experience, not just processing it. It's why a warehouse worker spots a glitch robots miss, or why an architect's wild sketch beats CAD's perfect lines.

This gap isn't a flaw to fix—it's a feature to harness. AI's best trick isn't autonomy; it's augmentation. Starbucks didn't axe baristas for robot brewers. In 2022, it rolled out AI to fine-tune inventory and staffing, cutting waste and wait times. Baristas got breathing room to chat with regulars, upsell pastries, build loyalty—the human stuff that drives profit. Sales climbed, not because machines took over, but because they cleared space for people to shine.

The Enhancement Edge: Real-World Proof

Let's widen the lens. In healthcare, AI's diagnostic chops are transforming clinics—but not by sidelining doctors. At Mount Sinai in New York, AI systems now screen mammograms, catching tumors radiologists might overlook. Yet the final call stays human. Why? Because patients don't just need a scan—they need a conversation, a plan, someone to trust. A 2023 trial showed AI-assisted doctors caught 20% more cancers than AI or humans alone. The combo wins, not the machine solo.

1 *The False Binary: Why AI Won't Replace Human Work*

In logistics, UPS leaned on AI to optimize delivery routes, slashing fuel costs by millions in 2022. Drivers didn't vanish—they adapted, handling quirks like gated estates or rush-hour snarls the algorithm couldn't predict. The tech shaved miles; humans kept it real. Creative fields follow suit. Pixar's artists use AI to render scenes at breakneck speed, but the storyboards—the heart of every film—stay hand-drawn by humans chasing a vision machines can't dream up.

Even language translation, where AI's made huge strides, leans on this dance. Google Translate can churn through 100 languages, but for legal contracts or poetry, human linguists step in. A 2023 study pitted AI against pros on French-to-English novels; the AI nailed grammar but flubbed idioms and tone. Humans caught the culture—AI just caught the words.

Reframing the Future

So why do we keep buying the replacement myth? Partly, it's the dazzle—AI's feats feel like magic, so we assume it's boundless. Partly, it's fear—disruption's real, and jobs *will* shift. But history whispers a steadier truth: tech amplifies us when we wield it right. The ATM didn't kill tellers; it multiplied branches. CAD didn't end architects; it fueled bolder buildings. AI won't erase work—it'll reshape it, pushing us toward what machines can't touch: judgment, empathy, imagination.

This isn't blind optimism. Disruption's coming—some roles, like data entry or rote coding, might shrink fast. But new ones will sprout, just as web design boomed after the internet hit. The trick is seeing AI as a partner, not a rival. Businesses chasing full automation might cut costs short-term, but the real winners—like Starbucks or Mount Sinai—are betting on enhancement. They're asking: *How do we supercharge our people?* Not: *How do we swap them out?*

For investors, this flips the script. Forget betting on AI to usurp humans—look for firms pairing tech with talent. For workers, it’s about leaning into what AI can’t do—solving the “what” while it handles the “how.” For policymakers, it’s crafting rules that keep humans in the driver’s seat, not the backseat.

The chapters ahead dig deeper. We’ll unpack AI’s guts (Chapter 2), spotlight what humans uniquely bring (Chapter 3), and map how this plays out across industries (Chapters 5–9). But the thread starts here: the future isn’t AI alone—it’s us, enhanced. That’s not a guess. It’s what the evidence, from ATMs to algorithms, keeps shouting.

The Current Thing

ChatGPT’s 2022 debut lit a match under AI hype. Suddenly, everyone had a taste—students cheating essays, coders auto-fixing bugs, execs dreaming of leaner payrolls. Stock tickers glowed green; pundits split between rapture and dread. But peel back the buzz, and the pattern’s familiar. At a Midwest ad agency, an AI tool churned out taglines in 2023—snappy, sure, but clients balked. They wanted ideas that *felt* human, not just sounded clever. The agency didn’t ditch copywriters; it paired them with AI to brainstorm faster, then refine with soul.

That’s the real story unfolding—not replacement, but retooling. AI’s rewriting the “how” of work, not the “why” or “what.” The firms getting it—like that agency, or UPS, or Pixar—aren’t automating people away. They’re amplifying what makes us irreplaceable. The rest of this book shows how.

2 Inside the Black Box: Understanding What AI Actually Does

Demystifying AI's real capabilities and limitations from an implementer's perspective

2.1 About AI

Artificial intelligence is a broad field which long-time researchers often jokingly define as “anything computers can’t do yet”. From early grammar checkers to chess to facial recognition, many features that are now routine were once considered AI. No doubt the same will eventually be said of the new generation of large language models (LLMs), the more precise term to describe the impressive new tools that include ChatGPT. Under the hood, LLMs are less magical and based on a straightforward application of an optimization algorithm called Generative Pre-trained Transformer (GPT) invented by Google researchers in 2017.

You can think of LLMs as a massively optimized and expanded version of the auto-complete feature your smartphone has featured for years. Instead of proposing the next word or two, LLMs can generate full sentences, paragraphs, books, and on

2 *Inside the Black Box: Understanding What AI Actually Does*

and on without limit. Its power comes from the GPT optimization that lets it take advantage of the massively-parallel architecture of graphic processing units (GPUs). Just as a graphical image can be broken into smaller pixels, each manipulated in parallel, LLMs break text documents into characters (or “tokens”) that are processed simultaneously within the GPU.

The GPT algorithm has one critical limitation: once set in motion, it cannot backtrack. Humans plan ahead, weigh different scenarios, and can change their minds based on foreseen alternatives. GPTs can only fake this planning ability through their access to mountains of data where such alternatives have already been explored. GPTs cannot do Sudoku, or handle chess boards not covered in its training books. Similarly, although it may *appear* to evaluate potential investment scenarios, it is merely spitting out a long stream of text that it harvested from options that were already evaluated somewhere in the bowels of its (massive) training sets.

It’s important to keep this “one-way” fact in mind when using LLMs. Because they have no concept of imagining how a future situation might change current plans, it would be wise to take its predictions with caution.

LLMs are *models* that compress all human knowledge – written, spoken, images, video – into a format that can generate similar-seeming content when given a starting prompt. Although the final models themselves are small enough to fit on a laptop or smartphone, they are created through a training process that consumes massive amounts of data — virtually everything on the public internet, plus collections of the text from millions of

books, magazines, academic journals, patent filings, and anything else its creators can find.

Thanks to a clever, time-saving shortcut discovered in the 2017 GPT algorithm, key parts of the training happen in parallel, limited only by the number of GPUs available. It's this optimization that explains the mad rush to buy GPUs, the chief beneficiary of which is Nvidia, thanks to its decades-long leadership in these fast processors. Although Nvidia chips were originally designed for fast graphics, their wide adoption means that many engineers are well-acquainted with CUDA, the low-level graphics programming software that powers Nvidia devices. When designing the various implementations of GPT, it was natural for developers to optimize for CUDA, further cementing Nvidia's lead.

Once trained, the LLM is a statistical prediction engine that knows the most likely word, phrase, or paragraphs that follow any given input. It knows, for example, that the phrase "Mary had a little" is highly likely to be followed by "lamb" or even the entire phrase "Its fleece was white as snow". It will apply the same statistical completion algorithm to any snippet of text, including those that look like questions, where the most likely "completion" is the answer to the question. For example, the statistically most likely way to complete the phrase "what is 1 + 1?" is "2".

The final LLM consists of billions of "parameters", finely-tuned statistical values created during the training process. But generating the response to your input requires similar levels of prodigious machine power. In fact, every character you type into the ChatGPT input box, as well as every character it types back, goes through many billions of computations. That slight delay you see as each character comes back at your terminal is not a clever UX effect intended to appear like a human is typing the answer.

In fact, the characters come out slowly because of the untold levels of computing power required to generate each one of them. Multiply this by the many millions of simultaneous ChatGPT users and you can understand why state-of-the-art LLMs are phenomenally expensive to operate.

Although these completions can be uncannily realistic, it's important to keep in mind that *it's just auto-completion*. Just as you would want to review an auto-complete suggestion before sending a reply on your smartphone, your ChatGPT answers require a similar level of skeptical scrutiny.

2.2 How does an LLM-based generative system work?

Imagine you have access to a zillion documents, preferably curated in some way reassures you about their quality and consistency. Wikipedia, for example, or maybe Reddit and other posts that have been sufficiently up-voted. Maybe you also have a corpus of published articles and books from trustworthy sources.

It would be straightforward to tag all words in these documents with labels like “noun”, “verb”, “proper noun”, etc. Of course there would be lots of tricky edge cases, but a generation of spelling and grammar-checkers makes the task doable.

Now instead of organizing the dictionary by parts of speech, imagine your words are tagged *semantically*. A word like “queen”, for example, is broken into the labels “female” and “monarch”; change the label “female” to “male” and you have “king”. A word like “Starbucks” might include labels like “coffee”, as well as “retail store” or even “Fortune 500 business”. You can shift the meaning by changing the labels.

Generating a good semantic model like this would itself be a significant undertaking, but people have been working on this for a while, and various good “unsupervised” means have been developed that can do this fairly well. For example, one trick might be to assign labels based on the types of words nearby. The word “Starbucks” means “Fortune 500 Business” if you find it in a paragraph containing words like “earnings” or “CEO”; but it means “coffee” if you see words like “\$4.95” or “latte”. This won’t be perfect, but you can imagine how it could get to be pretty good if you train on enough text.

2.3 Statistics of Words

This system works because words aren’t laid out randomly. Languages constrain the way words can follow one another. There are grammatical rules that determine word order, and there are additional semantic rules that further constrain which sentences make sense. “Colorless green ideas sleep furiously” is a grammatically valid sentence, but it makes no semantic sense and is extremely unlikely to occur. You and I know these rules because we’ve been living in the real world for many years. A computer can deduce most of these rules statistically as you give it more data.

The auto-completion feature, now ubiquitous in all word processors and mobile phones, is a simple application of the power of statistics combined with language. With any sized corpus, you’ll know with reasonable probability the likelihood that a particular word will follow another word. Interestingly, you can do this in any human language without even knowing about that language – the probabilities of word order come automatically from the sample sentences you have from that language.

2 Inside the Black Box: Understanding What AI Actually Does

Now go a step above auto-completion and allow for completion at the sentence level, or even the paragraphs or chapters. Given a large enough corpus of quality sentences, you could probably guess with greater-than-chance probability the kinds of sentences and paragraphs that should follow a given set of sentences. Of course it won't be perfect, but already you'd be getting an uncanny level of sophistication.

Pair this autocompletion capability with the work you've done with semantic labeling. And maybe go really big, and do this with even more meta-information you might have about each corpus. A Wikipedia entry, for example, knows that it's about a person or a place. You know which entries link to one another. You know the same about Reddit, and about web pages. With enough training, you could probably get the computer to easily classify a given paragraph into various categories: this piece is fiction, that one is medical, here's one that's from a biography, etc., etc.

Once you have a model of relationships that can identify the type of content, you can go the other direction: given a few snippets of one known form of content (biography, medical, etc.), "auto-complete" with more content of the same kind.

This is an extremely simplified summary of what's happening, but you can imagine how with some effort you could make this fairly sophisticated. In fact, at some level isn't that what we humans are already doing. If your teacher or boss asks you to write a report about something, you are taking everything you've seen previously about the subject and generating more of it, preferably in a pattern that fits what the teacher or boss is expecting.

Some people are very good at this: take what you heard from various other sources and summarize it into a new format.

“List five things wrong with this business plan”, you don’t necessarily need to *understand* the contents. If you’re good enough at re-applying patterns you’ve seen from similar projects, you’ll instinctively throw out a few tropes that have worked for you in the past. “The plan doesn’t say enough about the competition”, “the sales projections don’t take X and Y into account”, “How can you be sure you’ll be able to hire the right people”. There are thousands, maybe *hundreds of thousands* of books and articles that include these patterns, so you can imagine that with a little tuning a computer could get to be pretty good too.

2.4 Fine-tuning the output

Despite their impressive capabilities, raw language models often generate responses that are technically correct but practically useless. The key challenge lies not in generating text, but in producing output that aligns with human expectations and values.

Consider how a language model might handle the prompt “How do I make money quickly?” Without fine-tuning, it might suggest illegal schemes or unrealistic get-rich-quick schemes alongside legitimate business advice. The model is simply completing the text based on patterns it has seen, with no understanding of ethics or practicality.

This is where fine-tuning becomes crucial. The process involves three key approaches:

Reinforcement learning works by assigning scores to different outputs and adjusting the model’s parameters to favor higher-scoring responses. For example, responses suggesting legitimate business opportunities would receive higher scores

2 Inside the Black Box: Understanding What AI Actually Does

than those promoting scams. The model gradually learns to generate more appropriate suggestions.

RLHF (Reinforcement Learning with Human Feedback)¹ adds human judgment to this process. When developing ChatGPT, OpenAI presented human raters with multiple responses to the same prompt. Raters evaluated responses based on helpfulness, truthfulness, and potential harm. This feedback was then used to further train the model.

RLAIF (Reinforcement Learning with AI Feedback) attempts to scale this process by using AI systems to evaluate outputs. While more scalable than human feedback, this approach faces a fundamental challenge: how can an AI system evaluate responses without understanding human values and context?

The limitations of AI feedback become apparent in complex scenarios. Consider a medical chatbot trained using RLAIF. The AI evaluator might approve responses that are technically accurate but miss crucial context or nuance that a human doctor would catch. A response suggesting common flu treatments might be technically correct but dangerously incomplete if the patient has mentioned symptoms of a more serious condition.

This reveals a deeper issue with language models: they can't truly understand the implications of their outputs. They operate through pattern recognition rather than comprehension. When the model suggests "Take aspirin for your headache," it's not actually giving medical advice – it's predicting what words commonly follow "Take ____ for your headache" in its training data.

¹see Thomas Woodside and Helen Toner: How Developers Steer Language Model Outputs: Large Language Models Explained, Part 2 for a detailed but readable discussion.

2.4 Fine-tuning the output

Fine-tuning can mitigate these limitations but never fully eliminate them. Each approach comes with tradeoffs:

Traditional reinforcement learning excels at optimizing for specific, measurable metrics but struggles with subjective qualities like helpfulness or creativity. A model trained to maximize engagement might generate controversial or sensational content rather than accurate, balanced information.

RLHF produces more nuanced, socially aware responses but faces scaling challenges. Human evaluators are expensive and can introduce their own biases. Different evaluators might have conflicting views on what constitutes a “good” response, especially for controversial topics.

RLAIF attempts to bridge this gap but inherits the fundamental limitations of AI systems. An AI evaluator might approve responses that are syntactically perfect but semantically nonsensical, or miss subtle forms of bias that human evaluators would catch.

These challenges explain why even advanced language models occasionally produce responses that are simultaneously impressive and problematic. They might generate perfectly formatted code that contains subtle security vulnerabilities, or write eloquent essays that include factual inaccuracies.

Understanding these limitations is crucial for anyone implementing AI systems. The goal of fine-tuning isn’t to create perfect responses, but to make the model’s limitations predictable and manageable. This often means designing systems where AI augments rather than replaces human judgment.

Consider how JPMorgan uses AI in its risk assessment processes. Rather than relying solely on AI-generated analysis, they use fine-tuned models to flag potential issues for human

2 Inside the Black Box: Understanding What AI Actually Does

review. The AI processes vast amounts of data to identify patterns, but final decisions remain with human analysts who can consider broader context and implications.

This hybrid approach acknowledges both the power and limitations of fine-tuned models. The AI excels at pattern recognition and data processing, while humans provide judgment and context. The result is more effective than either could achieve alone.

2.5 Wisdom of the crowds

An LLM is sampling from an unimaginably complex mathematical model of the distribution of human words – essentially a wisdom of crowds effect that distills the collective output of humanity in a statistical way.

2.6 Where do you get the documents

OpenAI gets its documents from more than 200 million documents, 93% of which are in English, that are selected to be representative of a broad space of human knowledge.

Of course it starts with Wikipedia: almost 6 million articles.

One set of words comes from Common Crawl: a large, public-domain dataset of millions of web pages.

Another is a proprietary corpus called *WebText2* of more than 8 million documents made by scraping particularly high-quality

2.6 Where do you get the documents

web documents, such as those that are highly-upranked on Reddit.

Two proprietary datasets, known as *Books1* and *Books2* contain tens of thousands of published books. These datasets include classic literature, such as works by Shakespeare, Jane Austen, and Charles Dickens, as well as modern works of fiction and non-fiction, such as the Harry Potter series, The Da Vinci Code, and The Hunger Games.² There are also many other books on a variety of topics, including science, history, politics, and philosophy.

Also high on the list: b-ok.org No. 190, a notorious market for pirated e-books that has since been seized by the U.S. Justice Department. At least 27 other sites identified by the U.S. government as markets for piracy and counterfeits were present in the data set.

Washington Post has an interactive graphic that digs into more detail. (Also discussed on HN)

Yes, they crawl me:

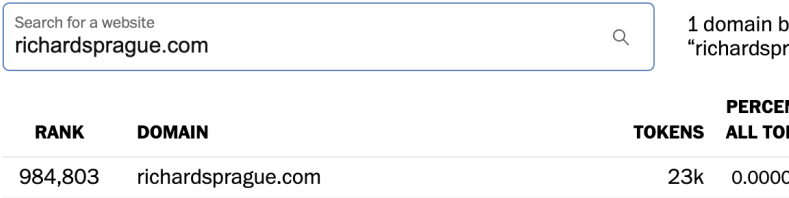
The websites in Google's C4 dataset

Search for a website blog.richardsprague.com		1 domain begins with "blog.richardsprague.com"	
RANK	DOMAIN	TOKENS	PERCENT OF ALL TOKENS
668,939	blog.richardsprague.com	35k	0.00002%

Figure 2.1: blog.richardsprague.com tokens on Google's C4 dataset

²see Apr 2023 Chang et al. (2023a)

The websites in Google’s C4 dataset



The Book believes it is important to present the complete contents of the data fed into

Figure 2.2: richardsprague.com tokens on Google’s C4 dataset

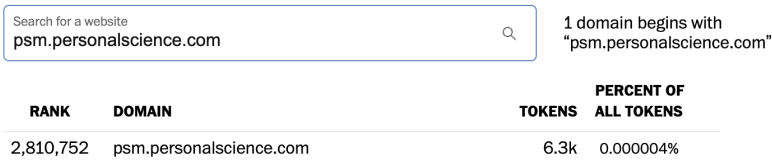


Figure 2.3: psm.personalscience.com tokens on Google’s C4 dataset

from NYTimes

2.7 Organizing the data

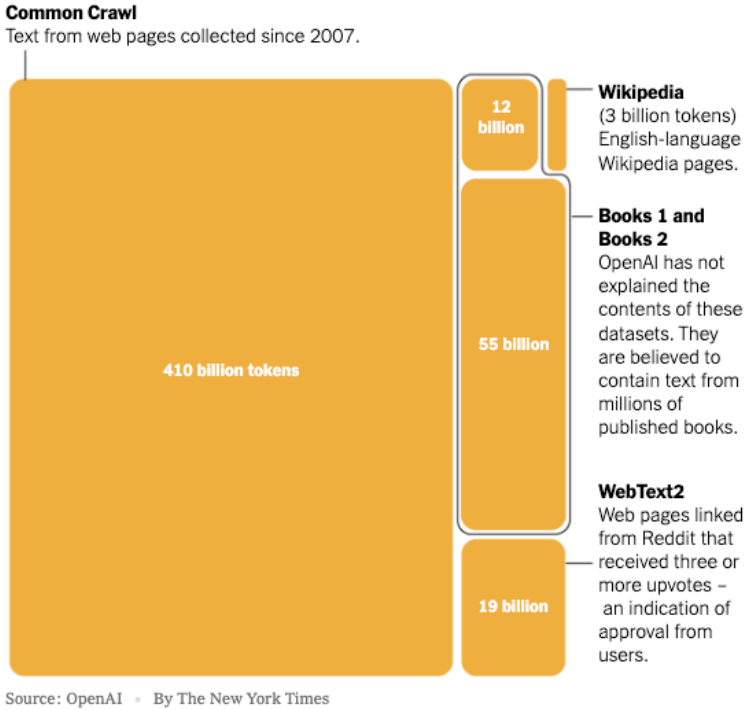


Figure 2.4: GPT-3 Data Sources

2.7 Organizing the data

At the core of a transformer model is the idea that many of the intellectual tasks we humans do involves taking one sequence of tokens – words, numbers, programming instructions, etc. – and converting them into another sequence. Translation from

2 Inside the Black Box: Understanding What AI Actually Does

one language to another is the classic case, but the insight at the heart of ChatGPT is that question-answering is another example. My question is a sequence of words and symbols like punctuation or numbers. If you append my question to, say, all the words in that huge OpenAI dataset, then you can “answer” my question by rearranging it along with some of the words in the dataset.

The technique of rearranging one sequence into another is called Seq2Seq*.

3 The What-How Divide

AI's Real Impact on Knowledge Work

Until recently, career success in knowledge work depended heavily on mastering “how” skills - knowing how to build a compelling PowerPoint, how to structure a financial model, or how to write efficient code. But as AI systems become more capable at these technical tasks, the competitive advantage is shifting dramatically toward people who know “what” needs to be done - those who can identify the right problems to solve and strategies to pursue.

This fundamental shift from “how” to “what” has profound implications for businesses, careers, and investment opportunities. Let's explore why this transformation is happening and what it means for different stakeholders.

3.1 The Traditional “How” Advantage

Traditionally, organizations needed large teams of specialists who knew “how” to perform various technical tasks: - Financial analysts who knew how to build complex Excel models -

3 The What-How Divide

Software engineers who knew how to write code in specific languages - Designers who knew how to use tools like Photoshop - Writers who knew how to craft clear technical documentation - Translators who knew how to convert text between languages

These specialists developed their skills through years of practice and training. Their expertise created both job security and earning power - companies were willing to pay premium salaries for people who could execute complex technical tasks effectively.

3.2 AI's Disruption of "How"

Large language models and other AI tools are rapidly getting better at many of these "how" tasks: - ChatGPT can write basic code in multiple languages - Midjourney can generate sophisticated images - Translation tools are approaching human-level quality - AI assistants can create presentations and documentation

This capability is expanding quickly. Tasks that seemed immune to automation just a few years ago are now being handled competently by AI systems. And unlike human specialists who may take years to master new skills, AI systems can be rapidly retrained or fine-tuned for new capabilities.

3.3 The Rise of "What" Skills

As AI handles more of the "how," competitive advantage shifts to people who excel at determining "what" needs to be done: - What problems are worth solving? - What features should a product include? - What markets should a company enter? -

3.4 Real-World Examples

What strategies will create sustainable advantages? - What metrics matter most for success?

These “what” decisions require capabilities that current AI systems fundamentally lack:

Pattern Recognition Across Domains Humans can notice subtle patterns and draw insights across seemingly unrelated fields. A business leader might see parallels between consumer behavior in fashion and trends in enterprise software, leading to novel strategic insights. Current AI systems, despite their broad training, struggle to make these creative connections in meaningful ways.

Judgment Under Uncertainty Many crucial business decisions involve incomplete information and conflicting priorities. Experienced leaders develop judgment about which risks are worth taking and which tradeoffs make sense. This type of judgment emerges from years of seeing both successes and failures firsthand - something AI systems cannot truly replicate.

Understanding Human Context Success in business ultimately depends on understanding human needs, motivations, and behaviors. While AI can process vast amounts of data about human behavior, it lacks the innate understanding that comes from being human and experiencing the full range of human emotions and social dynamics.

3.4 Real-World Examples

Let’s look at some specific examples of how this “what vs. how” divide plays out:

3.4.1 Book-Writing: When Bulldozers Move Words

What's the value of traditional books when ChatGPT can generate coherent answers to any question?

The analogy of construction work helps illustrate the relationship between AI and human authorship. Like bulldozers that efficiently move earth, AI can rapidly generate vast quantities of coherent text. But just as construction requires both heavy machinery and skilled artisans, meaningful books need both AI's raw productive power and human refinement.

Consider the process of learning chess. ChatGPT can explain rules, play practice games, and offer personalized instruction. Future versions might even customize the learning path based on individual aptitude and interests. However, a well-crafted book offers something different: a carefully structured approach that helps readers decide their level of engagement. The finite, constrained nature of a book provides focus that chatbots, with their endless potential for digression, cannot easily match.

The key to understanding AI's role in authorship lies in recognizing the distinct phases of book creation.

1. The initial phase - deciding subject matter and scope, aka the "what" phase — remains fundamentally human. While AI can help brainstorm ideas or identify underexplored topics, the essential creative spark and purpose must come from human intention. This reflects a broader truth about AI: it excels at processing existing patterns but struggles to generate truly novel directions.
2. The next phase — outlining the subject into smaller, related topics that make a coherent whole — demonstrates

3.4 Real-World Examples

the potential for human-AI collaboration. AI can quickly generate comprehensive topic structures, but human expertise is crucial for identifying gaps, inconsistencies, or areas requiring special emphasis. This interplay between AI's broad pattern recognition and human domain knowledge creates stronger frameworks than either could achieve alone.

3. The writing phase is where AI's "bulldozer" capabilities shine. Instead of laboriously crafting individual sentences, authors can use AI to generate substantial blocks of coherent text. This dramatically accelerates the initial draft process. However, like rough-graded earth, this AI-generated text requires careful refinement to achieve its final form.
4. The refinement phase is where human judgment becomes paramount. Authors must shape the AI-generated content to maintain consistent voice, ensure logical flow, and preserve the book's core purpose. This requires understanding nuances of audience expectations and subject matter that current AI systems cannot fully grasp.

This iterative process of generation and refinement continues until the project achieves its goals - another judgment that requires human evaluation. The result is neither purely AI-generated nor traditionally human-authored, but rather a new form of hybrid creativity that leverages the strengths of both.

The role of books may evolve, but their fundamental purpose - to present structured, focused exploration of subjects - will always be valuable. The challenge for authors is not to compete with AI's raw generative capabilities, but to use them effectively while maintaining the human elements that give books their lasting value.

3 The What-How Divide

This suggests a future where successful authors are those who master the art of AI collaboration rather than resist it. Just as modern architects must understand both traditional design principles and computer-aided tools, tomorrow's authors will need to balance classic writing skills with AI capabilities.

The key question is no longer whether AI will replace human authors, but how it will transform the authorship process. The answer lies in recognizing that while AI can move mountains of words, humans must still decide which mountains to move and how to shape the resulting landscape.

This transformation parallels broader changes in knowledge work. As AI handles more routine cognitive tasks, human value increasingly derives from higher-order skills like judgment, creativity, and strategic thinking. The future of authorship, like many professional fields, will belong to those who can effectively combine human insight with AI capabilities.

The rise of AI authors doesn't diminish the value of books but rather changes how they're created. The essential human elements - purpose, judgment, refinement - remain crucial, even as AI dramatically expands our capability to generate and process information. The result may be not just better books, but new forms of knowledge sharing that we're only beginning to imagine.

3.5 More examples

3.5.1 Software Development: Beyond Code Generation

The construction industry provides useful analogies for understanding AI's impact on software development. Just as modern

3.5 More examples

construction sites use both automated machinery and skilled human workers, software development is evolving into a hybrid process where AI handles routine coding tasks while humans focus on architecture and design decisions.

Consider a typical software project. Traditional development required writing every line of code manually, like building a house brick by brick. Now, AI coding assistants like GitHub Copilot or Amazon CodeWhisperer can generate entire functions or modules automatically, similar to how prefabricated components accelerated construction. These AI tools excel at producing standard elements - authentication systems, database queries, API endpoints - just as manufacturing automation excels at producing standardized building materials.

However, like construction projects, software development involves more than assembling standard components. A successful project requires understanding user needs, designing intuitive interfaces, ensuring security, and maintaining long-term reliability. These higher-level decisions remain firmly in human territory.

The architectural parallel is particularly apt. Just as architects must consider aesthetics, functionality, and structural integrity, software architects must balance user experience, system performance, and code maintainability. AI can suggest implementation details, but it cannot determine whether a feature aligns with business goals or how it might affect user behavior.

Technical debt offers another illuminating comparison. In construction, taking shortcuts (like using lower-grade materials) can speed completion but creates future maintenance problems. Similarly, in software development, quick fixes and temporary solutions accumulate as technical debt. While AI can identify potential debt and suggest refactoring strategies,

3 The What-How Divide

humans must weigh the business tradeoffs of addressing it now versus later.

Integration challenges further highlight AI's limitations. Modern software systems are complex ecosystems of interacting components, like cities with interconnected infrastructure systems. AI excels at optimizing individual components but struggles to understand system-wide implications. Humans must orchestrate these interactions, ensuring different parts work together coherently while maintaining system reliability and performance.

Security considerations demonstrate another crucial human role. Like building security systems, software security requires anticipating potential threats and implementing appropriate protections. AI can identify common vulnerabilities and suggest fixes, but it cannot understand the broader security context or evaluate risk tradeoffs. These decisions require human judgment informed by business context and threat assessment.

The testing and quality assurance phase reveals both AI's strengths and limitations. AI tools can automatically generate test cases and identify potential bugs, similar to automated building inspections. However, human testers are still essential for evaluating user experience, identifying edge cases, and ensuring the software meets business requirements. AI can verify that code works as written, but humans must verify it works as intended.

Looking ahead, successful software development will likely become increasingly collaborative between humans and AI. Development teams will need to master new workflows that leverage AI's capabilities while maintaining human oversight of critical

3.5 More examples

decisions. This might involve using AI for initial code generation and routine maintenance while focusing human effort on architecture, security, and user experience.

This evolution parallels broader trends in professional work. Just as power tools didn't eliminate the need for skilled carpenters but changed how they work, AI won't eliminate software developers but will transform their role. The most valuable developers will be those who can effectively direct AI tools while maintaining high-level system understanding.

The implications for software education and training are significant. Future developers will need less emphasis on memorizing syntax and more focus on system design, architecture, and AI collaboration skills. This mirrors how modern architectural education focuses less on manual drafting and more on design principles and computer-aided tools.

However, the fundamental role of human creativity and judgment remains unchanged. Just as beautiful buildings require human vision despite advanced construction technology, great software requires human insight despite sophisticated AI tools. The key is understanding AI as an enabler of human creativity rather than its replacement.

This suggests that software development is entering a new phase where success depends on effectively combining AI capabilities with human insight. The future belongs not to those who can code fastest, but to those who can best envision how technology can serve human needs while using AI to implement that vision efficiently and reliably.

In this new paradigm, the measure of a developer shifts from lines of code written to the effectiveness of their human-AI collaboration in creating valuable software solutions. The construction industry's evolution from manual labor to

3 The What-How Divide

machine-assisted craftsmanship provides a roadmap for this transformation.

3.5.2 Investment Analysis: Beyond the Numbers

Just as modern factories use automation for routine manufacturing while relying on human expertise for product design and quality control, investment analysis is evolving into a hybrid process where AI handles data processing while humans focus on strategic insights and judgment calls.

Consider a typical investment analysis project. Traditionally, analysts spent countless hours gathering financial data, creating comparison spreadsheets, and writing preliminary reports. Now, AI can instantly process quarterly reports, generate peer comparisons, and draft initial analyses. This is similar to how automated assembly lines handle routine manufacturing tasks, freeing human workers to focus on complex problems requiring judgment and creativity.

However, like manufacturing, successful investing involves more than processing standard inputs. While AI excels at identifying patterns in financial statements and market data, it struggles with crucial qualitative factors. Can management be trusted? Is the company's competitive advantage sustainable? Will current market opportunities persist? These questions require human judgment informed by experience and industry knowledge.

The manufacturing quality control parallel is particularly relevant. Just as experienced inspectors can spot subtle defects that automated systems miss, seasoned investors can identify red flags in management behavior or market dynamics that AI might overlook. A CEO's body language during earnings calls,

3.5 More examples

the timing of insider stock sales, or subtle shifts in competitive dynamics - these nuanced signals often prove more valuable than quantitative metrics.

Competitive analysis offers another illuminating comparison. In manufacturing, understanding market dynamics requires more than analyzing production statistics - it requires insight into changing consumer preferences, emerging technologies, and competitor strategies. Similarly, while AI can process vast amounts of market data, humans must evaluate whether a company's competitive position is truly defensible and whether management's strategy aligns with market realities.

The role of trust highlights another crucial human element. Just as manufacturing partnerships require trust built through personal relationships and demonstrated reliability, investment success often depends on accurately assessing management credibility. AI can flag inconsistencies in financial statements or unusual transaction patterns, but it cannot evaluate character or judge whether explanations for apparent irregularities are credible.

Market opportunity assessment demonstrates similar limitations. Like evaluating new manufacturing technologies, assessing market opportunities requires understanding both technical capabilities and human behavior. AI can analyze historical market data and identify trends, but it cannot predict how human customers, competitors, and regulators will react to new situations. These predictions require human insight into psychology and social dynamics.

Risk assessment reveals both AI's strengths and limitations. AI systems can quickly identify common risk factors and calculate standard metrics, similar to automated safety systems in manufacturing. However, the most significant risks often come from

3 The What-How Divide

unexpected directions that don't appear in historical data. Human judgment remains essential for identifying and evaluating these non-obvious risks.

Looking ahead, successful investment analysis will likely become increasingly collaborative between humans and AI. Analysis teams will need to master new workflows that leverage AI's data processing capabilities while maintaining human oversight of critical judgments. This might involve using AI for initial screening and routine monitoring while focusing human effort on qualitative assessment and strategic thinking.

This evolution parallels broader trends in professional work. Just as automation didn't eliminate the need for skilled manufacturing workers but changed their role, AI won't eliminate investment analysts but will transform how they work. The most valuable analysts will be those who can effectively direct AI tools while maintaining deep industry understanding and judgment capabilities.

The implications for investment education and training are significant. Future analysts will need less emphasis on spreadsheet skills and more focus on business judgment and AI collaboration capabilities. This mirrors how modern manufacturing education focuses less on manual skills and more on process management and technology integration.

However, the fundamental role of human judgment remains unchanged. Just as quality manufacturing requires human oversight despite advanced automation, successful investing requires human insight despite sophisticated AI tools. The key is understanding AI as an enhancer of human judgment rather than its replacement.

This suggests that investment analysis is entering a new phase where success depends on effectively combining AI capabilities

with human insight. The future belongs not to those who can process data fastest, but to those who can best understand business fundamentals while using AI to implement that understanding efficiently and reliably.

In this new paradigm, the measure of an analyst shifts from computational speed to the effectiveness of their human-AI collaboration in identifying truly attractive investments. The manufacturing industry's evolution from manual production to technology-enhanced craftsmanship provides a roadmap for this transformation.

3.5.3 AI and Healthcare: Beyond Pattern Recognition

The evolution of AI in healthcare parallels modern manufacturing quality control, where automated systems handle routine inspections while skilled technicians focus on complex problems requiring human judgment. Similarly, healthcare is becoming a hybrid system where AI processes medical data while human doctors focus on patient relationships and complex medical decisions.

Consider a typical diagnostic process. Traditionally, doctors spent considerable time reviewing test results, consulting medical literature, and documenting findings. Now, AI can instantly analyze lab results, medical images, and patient histories to suggest potential diagnoses. This is similar to how automated inspection systems quickly identify defects in manufactured products, allowing human inspectors to focus on more complex quality issues.

However, like quality control, successful healthcare involves more than pattern recognition. While AI excels at identifying anomalies in test results and suggesting standard treatments,

3 The What-How Divide

it struggles with crucial contextual factors. How will a patient's living situation affect treatment adherence? Which side effects are acceptable given a patient's lifestyle? What treatment modifications are needed given other health conditions? These questions require human judgment informed by direct patient interaction and medical experience.

The empathy factor is particularly relevant. Just as effective quality control requires understanding how products will be used in real-world conditions, effective healthcare requires understanding patients' lives and concerns. AI can process medical histories and suggest treatment protocols, but it cannot truly empathize with patient fears or understand how cultural and personal factors might affect treatment success.

Treatment customization offers another illuminating comparison. In manufacturing, standard quality metrics must often be adjusted for specific use cases. Similarly, while AI can recommend standard treatments based on medical literature, doctors must adapt these recommendations to individual patient circumstances. A treatment protocol that looks optimal on paper might be impractical or inappropriate given a patient's specific situation.

The trust relationship highlights another crucial human element. Just as manufacturing quality depends on trust between suppliers and customers, healthcare outcomes often depend on patient trust in their medical providers. AI can provide accurate medical information, but it cannot build the personal trust that encourages treatment compliance and honest symptom reporting.

Emergency response demonstrates both AI's strengths and limitations. AI systems can quickly process vital signs and suggest immediate interventions, similar to automated safety systems

3.5 More examples

in manufacturing. However, emergency medicine often requires split-second decisions based on incomplete information and complex tradeoffs. Human judgment remains essential for these high-stakes decisions where standard protocols may not apply.

Looking ahead, successful healthcare will likely become increasingly collaborative between humans and AI. Medical teams will need to master new workflows that leverage AI's analytical capabilities while maintaining human oversight of critical decisions. This might involve using AI for initial screening and routine monitoring while focusing human effort on patient interaction and complex case management.

This evolution parallels broader trends in professional work. Just as automation didn't eliminate the need for skilled quality control technicians but changed their role, AI won't eliminate doctors but will transform how they work. The most valuable healthcare providers will be those who can effectively direct AI tools while maintaining strong patient relationships and clinical judgment.

The implications for medical education and training are significant. Future doctors will need less emphasis on memorizing medical facts and more focus on patient communication and AI collaboration skills. This mirrors how modern quality control training focuses less on inspection procedures and more on system management and problem-solving.

However, the fundamental role of human judgment remains unchanged. Just as quality control requires human oversight despite advanced inspection technology, healthcare requires human insight despite sophisticated AI tools. The key is understanding AI as an enhancer of medical judgment rather than its replacement.

3 The What-How Divide

This suggests that healthcare is entering a new phase where success depends on effectively combining AI capabilities with human insight. The future belongs not to those who can recall the most medical facts, but to those who can best understand patient needs while using AI to implement that understanding efficiently and safely.

In this new paradigm, the measure of a healthcare provider shifts from diagnostic speed to the effectiveness of their human-AI collaboration in achieving optimal patient outcomes. The quality control industry's evolution from manual inspection to technology-enhanced oversight provides a roadmap for this transformation.

The challenge ahead is not whether to adopt AI in healthcare, but how to integrate it while preserving the human elements that make medicine effective. Success will require understanding both AI's capabilities and its limitations, while never losing sight of healthcare's fundamental mission: helping human patients achieve better health outcomes through personalized, compassionate care.

3.6 Investment Implications

This shift has important implications for investors:

Winners: - Companies that help humans make better “what” decisions - Tools that augment human judgment rather than replace it - Platforms that combine AI capabilities with human insight - Businesses with strong human judgment at their core

Losers: - Pure automation plays that don't preserve human judgment - Companies selling commoditized “how” skills - Businesses that can't articulate their human advantage

3.7 The Future of Work

This transition suggests several changes in how organizations will operate:

New Organizational Structures

- Flatter hierarchies as AI handles routine coordination
- Smaller, more senior teams focused on “what” decisions
- Greater emphasis on judgment and strategic thinking

Changed Skill Requirements

- Less focus on technical tool proficiency
- More emphasis on strategic thinking and judgment
- Greater value placed on cross-domain knowledge

Modified Training Approaches

- Reduced time spent teaching technical “how” skills
- Increased focus on judgment development
- More emphasis on understanding human factors

3.8 Preparing for the Transition

For individuals and organizations looking to succeed in this new environment, several approaches make sense:

For Individuals:

- Focus on developing judgment through varied experiences
- Build broad knowledge across multiple domains
- Practice making and learning from strategic decisions

3 The What-How Divide

- Get comfortable with ambiguity and uncertainty

For Organizations:

- Invest in tools that augment human judgment
- Develop processes that capture and share strategic insights
- Create cultures that value and develop good judgment
- Build teams with diverse perspectives and experiences

3.8.1 The Human Element Remains Central

It's crucial to remember that this shift doesn't diminish the importance of human contribution - it actually elevates it. As AI handles more routine tasks, human judgment, creativity, and wisdom become more valuable, not less.

Consider the example of chess: Despite AI systems being able to beat any human player, human chess hasn't disappeared. Instead, it's evolved. The most interesting matches now involve human-AI collaboration, where success depends on humans knowing what positions to play for and when to trust or override AI suggestions.

This pattern will likely repeat across many fields - the key to success will be understanding what humans do best and creating systems that augment these capabilities rather than try to replace them.

3.9 Looking Ahead

The transition from "how" to "what" won't happen overnight, but it's already underway. Organizations and individuals that

3.10 Beyond the False Binary

recognize and adapt to this shift will have significant advantages. Those that continue to focus primarily on “how” skills risk finding their capabilities increasingly commoditized by AI.

This shift also suggests we need to rethink education and training. Rather than focusing primarily on teaching technical skills that AI might soon handle, we should emphasize developing judgment, creativity, and strategic thinking - the fundamentally human capabilities that will become increasingly valuable.

The future belongs not to those who can execute tasks most efficiently, but to those who can best decide what tasks are worth doing in the first place.

In the early days of the personal computer revolution, spreadsheet software transformed financial analysis. Critics warned that tools like VisiCalc and Lotus 1-2-3 would eliminate financial analysts by automating their calculations. Instead, these tools dramatically increased productivity while shifting analysts’ focus from mathematical computation to business insight. Today’s artificial intelligence is driving a similar transformation, but at a far greater scale and across virtually every knowledge-based profession.

3.10 Beyond the False Binary

The current discourse on AI’s impact falls into a tiresome and inaccurate binary: either AI will replace human workers entirely, or its effects will be marginal. Both narratives miss the fundamental transformation underway. What we’re witnessing is not wholesale replacement but a profound shift in the nature of human contribution—a redistribution of value across the knowl-

3 The What-How Divide

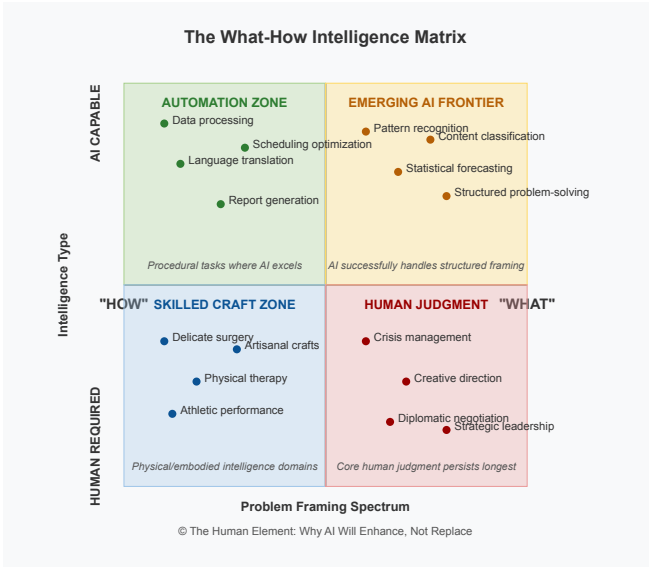


Figure 3.1

edge work spectrum that redefines which human capabilities command a premium.

This transformation becomes apparent when we distinguish between two fundamental aspects of any intellectual task: determining *what* needs to be done versus executing *how* to do it. This distinction, while seemingly straightforward, carries profound implications for the future of work, business strategy, and investment that extend far beyond the simplistic replacement narrative dominating public discourse.

The what-how framework offers remarkable clarity amid the confusing narratives surrounding AI. It helps explain why certain cognitive tasks are rapidly becoming commoditized while others remain stubbornly resistant to automation. More importantly, it provides a roadmap for individuals, organizations, and policymakers navigating a landscape where artificial intelligence increasingly pervades knowledge work.

3.11 The Nature of the Divide

Until the arrival of generative AI, individuals gained professional advantages through superior “how” skills—they excelled at crafting compelling presentations, building complex spreadsheets, writing efficient code, or translating between languages. These implementation abilities represented valuable skills that could be developed and applied on behalf of those who determined *what* needed to be done.

In traditional organizational hierarchies, executives and managers typically decide *what* initiatives to pursue, while specialized knowledge workers determine *how* to execute them. This division has historically functioned efficiently because

3 *The What-How Divide*

how expertise—whether in financial modeling, software development, or content creation—required significant investment in learning specialized tools and methodologies.

The emergence of sophisticated AI systems fundamentally alters this equation. Large language models demonstrate remarkable proficiency in implementation tasks, often exceeding human capabilities in narrow domains. They can generate code, compose business communications, create visual assets, and perform complex analyses with minimal human guidance. These systems excel precisely in the domain of *how*—the execution of well-defined tasks within established parameters.

What these systems cannot do—and what remains uniquely human—is determine *what* is worth doing in the first place. They cannot independently identify which problems merit attention, which strategies align with organizational values, or which approaches will resonate with stakeholders. They lack the contextual understanding, ethical framework, and strategic vision required to make these determinations.

Consider the financial analyst whose value traditionally derived from technical modeling skills. As AI systems increasingly automate complex financial calculations, the analyst's competitive advantage shifts toward identifying which factors merit analysis, which comparisons yield strategic insights, and how findings translate into investment decisions. The technical implementation—the *how*—becomes commoditized, while judgment about *what* to analyze becomes the primary value driver.

This pattern repeats across knowledge work domains. In marketing, AI can generate endless variations of campaign materials, but cannot determine which messaging will align with brand

3.12 *Philosophical Dimensions of the Divide*

values and audience expectations. In software development, AI can produce functional code based on specifications but cannot identify which features will deliver genuine user value. In healthcare, AI can analyze diagnostic images with remarkable accuracy but cannot integrate these findings with the full context of patient well-being.

3.12 Philosophical Dimensions of the Divide

The what-how divide resonates with deeper philosophical questions about the nature of intelligence and agency. Martin Heidegger, whose work we explore more fully in Chapter 4, offers particularly relevant insights through his concept of “comportments”—the way humans face and engage with the world around them.

When we are deeply engaged in an activity—skillfully driving a car, playing an instrument, or writing code—we are not consciously thinking about the mechanics of these actions. Our focus extends beyond the immediate task to its purpose and meaning within our broader existence. The ultimate comportment, Heidegger suggests, is our orientation toward being itself, which encompasses our understanding of past, present, and future.

Artificial intelligence systems, even sophisticated ones like GPT-4 or Claude, lack these comportments. They process information without any inherent purpose or temporal orientation. They can mimic human-like outputs but have no concept of why these outputs matter or how they fit into broader human concerns. This philosophical distinction manifests practically in AI’s inability to determine *what* is worth doing independent of human direction.

3 *The What-How Divide*

The what-how divide thus represents more than a practical delineation of tasks; it reflects a fundamental distinction between human and artificial intelligence. While AI excels at executing well-defined processes—the *how*—it cannot engage with the existential questions of purpose and meaning that inform human decisions about *what* deserves attention.

This philosophical perspective helps explain why LLMs struggle with certain seemingly simple tasks, as we demonstrated in prior chapters. Tasks that require constant re-evaluation and adjustment based on evolving goals—like writing a sentence that accurately describes its own length or completing a Sudoku puzzle—reveal the fundamental limitations of systems that cannot backtrack or reconsider their approach once they’ve begun generating outputs.

3.13 Case Studies: The Divide in Practice

To illustrate the what-how divide, let’s examine several domains where this transformation is particularly evident:

3.13.1 Software Development

Traditional programming expertise focused heavily on implementation details—mastering specific languages, frameworks, and architectural patterns. While these technical skills remain valuable, AI code generation tools increasingly automate routine implementation tasks. The premium shifts toward determining which features will deliver value, how systems should interact with users, and what architectural decisions will support long-term business objectives.

3.13 Case Studies: The Divide in Practice

Senior developers report that junior programmers who once spent years mastering syntax and debugging techniques now leverage AI assistants to handle these aspects, allowing them to focus earlier in their careers on higher-level system design and user experience considerations—traditionally the domain of more experienced developers.

This shift alters the career progression trajectory for software engineers. Technical implementation skills remain necessary but insufficient; they must be paired with strategic judgment about what deserves implementation in the first place. Engineers who maintain purely technical focus without developing this broader perspective may find their competitive position eroding as AI systems increasingly automate routine coding tasks.

3.13.2 Content Creation

In media and marketing, AI systems now generate remarkably coherent and stylistically appropriate content at scale. The limiting factor is no longer production capacity but strategic direction—determining which messages will resonate with target audiences, which topics deserve attention, and how content aligns with broader brand narratives.

Marketing executives evaluating AI writing assistants frequently report that while these tools can “automatically compose email replies,” users typically spend as much time editing these drafts as they would creating responses from scratch. The real value emerges when humans with deep customer knowledge direct these tools toward specific strategic objectives.

3 *The What-How Divide*

This transformation extends beyond business communication to creative fields. As we explored with the AI-generated completion of Beethoven's unfinished tenth symphony, technical proficiency alone cannot replicate the ineffable quality that distinguishes truly meaningful creative work. As music critic Jan Swafford observed, "We humans need to see the human doing it." The value derives not just from the output itself but from knowing it represents authentic human struggle, insight, and purpose.

3.13.3 Healthcare

Medical diagnostic systems increasingly match or exceed human performance in analyzing medical images, identifying patterns in patient data, and suggesting potential diagnoses. Yet these systems cannot determine which factors are most relevant for a particular patient, how to weigh complex trade-offs between treatment options, or how to communicate findings in ways that respect patient values and preferences.

A physician whose only skill is knowing *how* to diagnose a patient's condition is becoming less necessary. The crucial human contribution shifts toward determining *what* aspects of patient wellbeing deserve priority, which treatment approaches align with patient values, and how to integrate medical insights with broader quality-of-life considerations.

This shift carries significant implications for medical education and practice. Technical diagnostic skills remain essential but must increasingly be paired with heightened capabilities for integrative judgment, ethical reasoning, and communication. The most effective healthcare practitioners of the future will leverage AI for routine analytical tasks while focusing their

human expertise on the complex judgments that machines cannot make.

3.14 The Competitive Dynamics of the Divide

The what-how framework carries significant implications for competitive strategy across industries. As implementation capabilities become increasingly commoditized through AI, sustainable competitive advantage shifts toward superior judgment about what deserves implementation in the first place.

This dynamic particularly challenges organizations that have traditionally derived their advantage primarily from superior execution. When AI systems can implement strategies with comparable efficiency across competitors, the primary differentiator becomes the quality of strategic judgment guiding that implementation. Organizations must evolve their capabilities accordingly, developing institutional capacity for the complex judgments that remain resistant to automation.

We see this pattern emerging in investment management, where quantitative analysis tools have become increasingly sophisticated and widely available. The differentiator for successful investment firms shifts toward superior judgment about which factors merit analysis, which market signals deserve attention, and how various considerations should be weighted in decision-making.

Similarly, in management consulting, the technical aspects of data analysis and presentation—traditionally key components of the service offering—are increasingly automated. The value

3 *The What-How Divide*

proposition shifts toward helping clients determine which problems deserve attention, which approaches align with organizational values, and how various factors should be prioritized.

For technology companies specifically, the what-how framework offers valuable guidance for product development. The most successful AI implementations enhance rather than replace human judgment, allowing people to focus on the high-value *what* decisions where they maintain a durable advantage. Products that merely automate implementation without facilitating better strategic decisions will struggle to deliver sustainable value.

3.15 Organizational Implications

This fundamental shift carries significant implications for how organizations approach talent development, operational structure, and competitive strategy. Companies that recognize and adapt to the what-how divide will establish sustainable advantages in an AI-enhanced economy.

First, talent development programs must evolve beyond technical training focused on implementation skills. While baseline technical literacy remains essential, organizations should invest more heavily in developing employees' abilities to frame problems effectively, synthesize insights across domains, and make nuanced judgments that integrate technical, business, and ethical considerations.

The most valuable professional development initiatives will foster precisely those capabilities that remain distinctly human—contextual understanding, strategic synthesis, and ethical judgment. This represents a significant departure from

3.15 Organizational Implications

traditional approaches that emphasize mastery of specific tools and methodologies.

Second, workflow design should consciously separate strategic decisions from implementation details, creating clear interfaces between human judgment and AI execution. This approach maintains appropriate human oversight while leveraging AI's capabilities for rapid, consistent implementation.

Effective workflow design requires careful consideration of where human judgment adds the most value. Rather than automating entire processes end-to-end, organizations should identify the critical decision points where human judgment remains essential and design workflows that explicitly incorporate this judgment while automating surrounding implementation steps.

Third, organizational structures should evolve to emphasize roles that combine domain expertise with AI literacy. The traditional separation between business strategists and technical implementers becomes less valuable as AI systems increasingly bridge this gap. New hybrid roles will emerge that focus on translating business objectives into effective AI implementation approaches.

This structural evolution may require reconsidering traditional career paths and reporting relationships. Organizations that maintain rigid distinctions between technical and strategic roles may struggle to develop the integrated capabilities needed for effective human-AI collaboration.

3.16 Investment Implications

For investors, the what-how framework offers valuable guidance for evaluating AI-related opportunities. Companies positioned to win in this environment include those that:

1. Develop tools that enhance human strategic thinking rather than merely automating implementation tasks
2. Create platforms that facilitate seamless collaboration between human judgment and AI execution
3. Build solutions that maintain appropriate human oversight while leveraging AI capabilities
4. Design business models that recognize and reward uniquely human contributions

By contrast, companies that focus exclusively on automation without considering the continued importance of human judgment will likely struggle to deliver sustainable value. The most successful AI implementations will be those that augment rather than replace human capabilities, allowing people to focus on the high-value *what* decisions where they maintain a durable advantage.

This perspective offers a useful corrective to the common investor tendency to overvalue pure automation plays. The history of technology adoption suggests that approaches that enhance rather than replace human capabilities typically deliver more sustainable value over time.

3.17 The Evolution of Knowledge Work

As AI capabilities continue to evolve, we can anticipate further shifts in the relative value of different forms of human contribution. Implementation skills—the *how*—will continue to be commoditized, while strategic judgment—the *what*—will command an increasing premium. This doesn't mean implementation expertise becomes irrelevant, but rather that it must be paired with higher-level strategic capabilities to remain valuable.

For individual knowledge workers, this suggests a clear direction for professional development. Rather than focusing exclusively on technical mastery within narrow domains, sustainable career advancement will require developing broader strategic capabilities: understanding stakeholder needs, synthesizing insights across disciplines, and making nuanced judgments that integrate technical, business, and ethical considerations.

For educational institutions, the what-how divide suggests the need for fundamental curriculum redesign. Traditional education systems heavily emphasize *how* skills—teaching specific methodologies, tools, and techniques. Future-oriented education should place greater emphasis on developing students' abilities to frame problems effectively, think across disciplinary boundaries, and make contextual judgments that cannot be easily automated.

For policymakers, this framework offers a more nuanced understanding of AI's impact on employment and economic opportunity. Rather than focusing exclusively on potential job displacement, policy approaches should consider how to facilitate the transition toward work that emphasizes uniquely human strategic capabilities while ensuring that the benefits of AI-driven productivity gains are broadly shared.

3.18 Integration with Enhancement Thesis

The what-how framework aligns perfectly with our core thesis that AI will enhance rather than replace human capabilities across industries. By automating routine implementation tasks, AI frees human cognitive capacity for higher-level strategic thinking—the domain where human judgment maintains a durable advantage. This represents not replacement but enhancement of human potential.

This perspective also explains why purely automated approaches often disappoint. When AI systems operate without appropriate human direction and oversight, they may execute flawlessly within their parameters while completely missing the broader context that gives their outputs meaning and value. The most successful implementations maintain humans “in the loop” precisely because human judgment about *what* matters cannot be delegated to automated systems.

Consider full self-driving technology, which we’ll explore more fully in later chapters. Companies like Tesla have collected unprecedented amounts of driving data and developed increasingly sophisticated systems for navigating complex environments. Yet as robotics pioneer Rodney Brooks has observed, these systems still struggle with the contextual judgment that experienced human drivers exercise effortlessly.

A human driver approaching a neighborhood with cars parked tightly on both sides naturally slows down, recognizing the increased risk of children darting into the street. This judgment doesn’t derive from explicit rules but from a holistic understanding of context that integrates multiple factors—some explicit, others tacit. Autonomous systems may eventually

3.19 Conclusion: Navigating the Transformation

replicate this behavior through sophisticated pattern recognition, but they cannot independently determine which factors deserve attention without human direction.

3.19 Conclusion: Navigating the Transformation

The what-how divide provides a powerful framework for understanding AI's true impact on knowledge work and business strategy. Rather than wholesale replacement, we're witnessing a fundamental shift in the nature of human contribution—from executing well-defined tasks to making strategic judgments about what deserves attention and how different considerations should be weighed.

This transformation presents both challenges and opportunities. Organizations and individuals that continue to focus exclusively on implementation skills will find their competitive position eroding as AI systems increasingly automate these functions. Those who develop the strategic judgment to determine *what* is worth doing—and the ability to direct AI systems effectively toward these ends—will thrive in an AI-enhanced economy.

The what-how framework aligns with our broader thesis that successful AI implementation requires keeping humans “in the loop.” Not because of temporary technical limitations that will eventually be overcome, but because of fundamental differences between human and artificial intelligence. The most valuable human contributions have always involved more than technical execution—they reflect purpose, meaning, and judgment that remain uniquely human even as AI capabilities advance.

3 The What-How Divide

In the next chapter, we'll explore these philosophical dimensions more deeply, examining why understanding the nature of human intelligence is crucial for designing effective human-AI collaborations. By recognizing both the capabilities and limitations of artificial intelligence, we can develop approaches that truly enhance human potential rather than attempting to replace it.

4 Beyond Computation: The Philosophy of Human Intelligence

What Heidegger and other thinkers reveal about the fundamental differences between human and artificial intelligence

Before we get started, I have a few words to say....

Previous chapters examined AI's capabilities and limitations from technical and business perspectives. But to truly understand why human intelligence remains irreplaceable, we need to dig deeper into what makes human thinking unique. This takes us into philosophical territory that might seem abstract at first but has profound practical implications for business leaders and investors trying to navigate the AI revolution.

While American and British philosophers have focused primarily on logic and language – which certainly matter for AI development – Continental philosophers, particularly Martin Heidegger, tackled more fundamental questions about what it means to think and exist. Their insights help explain why even our most advanced AI systems, despite impressive capabilities, still miss essential aspects of human intelligence.

4 Beyond Computation: The Philosophy of Human Intelligence

The fundamental issue is that we've inherited a flawed model of human intelligence from Descartes and other early modern philosophers. They viewed humans as essentially thinking machines – hence “I think, therefore I am.” This same assumption underlies most AI development: if we can replicate human-like information processing, we'll achieve human-like intelligence. But this gets things exactly backwards.

We're not primarily thinking machines that sometimes act in the world. Instead, we're fundamentally beings-in-the-world (Heidegger's hyphenated term emphasizes this unity) who sometimes step back to think abstractly. This distinction has enormous implications for how we should think about AI and its limitations.

Consider a skilled trader on a busy trading floor. When they're “in the zone,” they're not consciously thinking through each decision. They're responding to market movements, news flows, and subtle signals from colleagues with an intuitive grasp that comes from years of embodied experience. Heidegger would say they're exhibiting “ready-to-hand” engagement with their environment, not detached analytical thinking.

This is fundamentally different from how AI trading systems work. The AI processes data and applies algorithms, but it lacks what Heidegger calls “comportment” – that basic way of being oriented toward and engaged with the world that comes before any explicit thinking. This explains why pure algorithmic trading works well for certain types of high-frequency operations, but the most successful hedge funds still rely heavily on human judgment for their major positions. The humans aren't necessarily “smarter” than the algorithms – they just engage with the market in a fundamentally different way.

This connects to another key Heideggerian insight: we're

temporal beings who inherently understand past, present, and future as a unified whole. When a skilled investor or business leader makes decisions, they're not just processing current data – they're drawing on their lived experience of the past and projecting possibilities into the future. AI systems, in contrast, can only process historical data and make statistical projections. They lack what Heidegger calls “temporality” – that basic human way of existing across time that makes genuine understanding possible.

This explains why many business leaders discover that their best human decision-makers aren't just processing more data – they're bringing something qualitatively different to the table. Humans don't primarily understand things by building up from basic facts to complex conclusions. Instead, we always already have what Heidegger calls a “pre-understanding” – a practical grasp of how things work that comes from being embedded in a shared world of meaning.

Think about how a seasoned executive “reads the room” in a crucial negotiation. They're not just processing verbal statements and body language signals. They're drawing on a lifetime of cultural and social understanding that no AI system can replicate because AIs lack what Heidegger calls “being-with” – that fundamental way humans share a meaningful world with others.

This explains something often observed in investment teams: Junior analysts may have impressive technical skills and can process more data than their senior colleagues. But the best senior investors have something that can't be reduced to information processing – a kind of practical wisdom that comes from years of being immersed in markets and business.

These philosophical insights have practical implications for how

4 Beyond Computation: The Philosophy of Human Intelligence

businesses should implement AI. Consider three different business activities:

1. Processing insurance claims
2. Negotiating a major acquisition
3. Developing a new product strategy

The first task is mainly about following procedures and processing information – perfect for AI enhancement. The second requires deep cultural understanding and reading subtle human dynamics – AI can assist but human judgment remains essential. The third requires what Heidegger calls “projection” – understanding current possibilities in light of future potential. AI can provide data and analysis, but only humans can truly innovate because only humans exist temporally.

This pattern appears consistently in markets. Companies that try to completely automate complex human judgments often disappoint, while those that use AI to enhance human capabilities tend to succeed. It’s not about replacing human intelligence but augmenting it in ways that respect its unique character.

This suggests the current focus on making AI more “human-like” may be misguided. Instead of trying to replicate human intelligence, which is fundamentally embedded in being-in-the-world, we should focus on developing AI systems that complement human capabilities. Think about how a hammer extends human capabilities without trying to replicate the human arm. Similarly, AI should extend human intelligence without trying to replicate human understanding.

For investors, this means companies that understand these distinctions – between what AI can enhance and what remains irreducibly human – are more likely to successfully implement AI

than those pursuing full automation of human judgment. It also suggests we need to rethink how we evaluate AI progress. Instead of asking whether AI can pass increasingly sophisticated Turing tests, we should ask how effectively it enhances distinctively human capabilities.

The goal shouldn't be artificial general intelligence that replicates human thinking. Instead, we should aim for artificial specific intelligence that amplifies human judgment while respecting its unique character. This philosophical perspective helps explain why the most successful AI implementations are those that enhance rather than replace human judgment. They succeed not despite keeping humans in the loop, but because they maintain that crucial human element.

This brings us back to our core enhancement thesis. By understanding the fundamental differences between human and artificial intelligence, we can better appreciate why enhancement rather than replacement is the right goal. The future belongs not to pure AI systems, but to human-AI partnerships that respect and amplify what makes human intelligence unique – our being-in-the-world, our temporality, and our fundamental way of sharing meaning with others.

These insights have profound implications for how businesses should approach AI implementation, which we'll explore in the following chapters. But the key takeaway is this: successful AI strategy requires understanding not just what computers can do, but what makes human intelligence irreplaceably unique.

5 The Human Edge

Judgment in an Age of Algorithms

In an era increasingly defined by algorithmic processing, the question of human judgment's unique value becomes not merely philosophical but practical. The rapid advancement of artificial intelligence has created a peculiar paradox: as machines become more capable of executing sophisticated tasks, the most distinctly human capacities become more valuable, not less. To understand this paradox requires careful examination of what constitutes judgment and why it remains stubbornly resistant to computational replication.

5.1 The Uniqueness of Human Judgment

Consider the ambitious attempt to create Beethoven's unfinished tenth symphony using artificial intelligence. The project, undertaken by Playform AI, represented a perfect test case for understanding the boundaries between algorithmic production and human creation. The team trained sophisticated models on Beethoven's complete works, incorporating fragments and sketches the composer had left for his tenth symphony. The result was technically proficient—notes arranged in patterns

5 *The Human Edge*

statistically consistent with Beethoven's compositional style. Yet something essential was missing.

Music critic Jan Swafford's assessment was unequivocal: "aimless and uninspired." What Swafford identified was not merely technical deficiency but the absence of struggle, refinement, and contextual understanding that characterized Beethoven's actual creative process. The composer's drafts were often mundane until transformed through iterative revision guided by judgment—a quality that emerges from being situated in a cultural, historical, and emotional context that no algorithm, however sophisticated, currently inhabits.

This observation extends beyond music. Across domains—from sports to business leadership, from medical diagnosis to strategic planning—we find consistent evidence that human judgment operates differently from algorithmic processing. The difference lies not merely in computational capacity but in the nature of understanding itself.

Martin Heidegger's philosophical framework provides valuable insight here. Heidegger challenged the Cartesian notion that human intelligence is primarily computational, arguing instead that our fundamental relationship with the world is one of "being-in-the-world" (*Dasein*). From this perspective, understanding emerges not from abstract calculation but from practical engagement with a meaningful context. Humans do not process the world as detached observers calculating optimal responses; rather, we inhabit it as participants whose very perception is structured by practical concerns and possibilities.

When we navigate complex situations—whether negotiating a business deal, diagnosing an unusual medical condition, or responding to unexpected market shifts—we draw upon this em-

bodied understanding. We recognize patterns not as statistical correlations but as meaningful constellations of relevance. This capacity for situated judgment represents what philosopher Hubert Dreyfus, interpreting Heidegger, called “comportment”—an orientation toward the world that precedes and enables explicit reasoning.

Artificial intelligence systems, while increasingly sophisticated in their pattern recognition capabilities, operate fundamentally differently. They recognize statistical regularities without inhabiting the human world of concerns and commitments. This distinction becomes apparent when examining the architecture of both human and algorithmic judgment.

5.2 The Architecture of Judgment

Human judgment integrates multiple dimensions of understanding that current AI systems struggle to replicate. Consider the limitations of large language models (LLMs), which represent state-of-the-art capabilities in natural language processing. These systems excel at pattern recognition and statistical inference but encounter fundamental limitations when faced with tasks requiring genuine understanding.

The inability of LLMs to “backtrack”—to revise fundamental assumptions mid-stream—represents more than a technical limitation. It reveals a structural difference between statistical pattern completion and genuine understanding. When humans engage in complex reasoning, we constantly revise our approach based on emerging information, testing alternative frames of reference and adjusting our conceptual foundations. This capacity for recursive self-correction reflects our temporality—our ability to hold past, present, and future in dynamic tension.

5 *The Human Edge*

For example, when confronted with the task of writing “a sentence that describes its own length in words,” LLMs consistently fail despite their impressive capabilities. The task requires not merely statistical inference but meta-cognitive awareness—the ability to simultaneously generate content while monitoring and adjusting that content against an evolving standard. This capacity for self-reference and dynamic adjustment characterizes human judgment across domains.

Equally significant is what philosopher Michael Polanyi termed “tacit knowledge”—understanding that cannot be fully articulated in explicit terms. Expert clinicians recognize patterns of disease before they can articulate the specific indicators that triggered their concern. Experienced investors sense market shifts through subtle cues that precede formal indicators. This dimension of understanding emerges from embodied experience accumulated over years of immersion in particular contexts.

The distinction parallels what we might call the “what-how” divide in contemporary knowledge work. Artificial intelligence excels at executing “how” tasks—implementing specific procedures once objectives have been defined. The increasing capability of AI systems to execute these procedural tasks generates enormous efficiency gains across industries. Yet these gains simultaneously increase the premium on “what” intelligence—the capacity to determine meaningful objectives, frame problems effectively, and identify relevant contexts for analysis.

5.3 The “What-How” Divide in Professional Contexts

Financial markets provide a particularly instructive domain for examining this distinction. Quantitative models have transformed investment management, enabling sophisticated analysis of vast datasets and revealing patterns invisible to unaided human perception. Yet the most successful investment approaches typically integrate algorithmic analysis with human judgment rather than replacing the latter with the former.

This integration recognizes that market behavior reflects not merely mathematical relationships but complex human psychology, institutional dynamics, and contextual factors that resist complete formalization. The 2008 financial crisis illustrated the dangers of excessive reliance on quantitative models that failed to account for human behavior under exceptional conditions. Similarly, the unprecedented monetary interventions following the COVID-19 pandemic created market conditions that defied historical patterns, requiring judgment to navigate effectively.

The most sophisticated hedge funds and investment firms have therefore developed what might be termed “judgment architectures”—organizational structures that integrate algorithmic processing with human expertise. These architectures recognize that algorithms excel at processing vast datasets and identifying statistical patterns, while human judgment excels at integrating these patterns with broader contextual understanding and adapting to novel situations.

Similar patterns emerge in technical implementation across industries. Consider the development of fully autonomous

5 The Human Edge

vehicles, which represents one of the most ambitious applications of artificial intelligence to real-world problems. Despite massive investments and impressive technical achievements, full autonomy remains elusive in complex, unpredictable environments.

The challenges facing autonomous vehicle systems reveal the limitations of purely algorithmic approaches to navigation and decision-making. While these systems excel at processing sensor data and executing well-defined maneuvers, they struggle with the contextual understanding that human drivers develop through embodied experience. A human driver intuitively recognizes that children playing near a street require extra caution, that an unusually positioned vehicle might indicate an unseen hazard, or that specific weather conditions might affect road surfaces in ways not immediately visible.

Rodney Brooks, robotics pioneer and former director of MIT's Computer Science and Artificial Intelligence Laboratory, has consistently emphasized these limitations. His predictions regarding autonomous vehicle development have proven remarkably accurate, with full autonomy consistently arriving later than industry projections. Brooks understands that navigating physical environments requires not merely sophisticated sensors and algorithms but contextual understanding that emerges from being situated in a meaningful world.

5.4 Decision-Making Under Uncertainty

Perhaps the most significant advantage of human judgment becomes apparent under conditions of genuine uncertainty. Algorithmic approaches excel at optimizing decisions under

5.4 Decision-Making Under Uncertainty

risk—situations where potential outcomes and their probabilities can be reasonably estimated. They struggle, however, with uncertainty—situations involving unknown variables, emergent phenomena, and fundamental indeterminacy.

This distinction becomes particularly relevant in domains characterized by complexity, path dependency, and human interaction. Consider pandemic response planning, where initial frameworks must adapt to evolving viral behavior, social dynamics, and institutional constraints. The COVID-19 pandemic revealed both the value of algorithmic modeling and its limitations when confronting genuinely novel situations. The most effective responses integrated computational modeling with expert judgment that could adapt to emerging information and contextual factors.

The limitations of purely algorithmic approaches under uncertainty relate to what we might call the “paradox of explicability.” Organizations increasingly demand explainable AI—systems whose recommendations can be traced to transparent reasoning processes. Yet humans routinely trust human experts whose intuitive judgments cannot be fully articulated. We accept that an experienced physician’s concern might precede explicit justification or that a seasoned investor’s caution might reflect pattern recognition too subtle for immediate expression.

This asymmetric standard reflects an implicit understanding that human judgment operates differently from algorithmic processing. We recognize that human experts integrate explicit knowledge with tacit understanding developed through situated experience. This integration enables what philosopher Charles Sanders Peirce termed “abduction”—the generation of novel hypotheses that cannot be derived through purely deductive or inductive reasoning.

The capacity for abductive reasoning becomes particularly valuable when confronting black swan events—high-impact developments that lie outside normal expectations and resist prediction through historical analysis. The financial market disruptions following the 2001 terrorist attacks, the 2008 financial crisis, and the COVID-19 pandemic each required judgment that could transcend historical patterns and recognize emergent possibilities.

5.5 The Enhancement Framework Revisited

Understanding these distinctive capacities allows us to develop more effective approaches to human-AI collaboration. Rather than conceptualizing artificial intelligence as a replacement for human judgment, we can design systems that enhance human capabilities by performing complementary functions. This enhancement framework acknowledges the distinctive strengths of both human judgment and algorithmic processing.

Effective enhancement requires careful attention to interface design, workflow integration, and organizational architecture. Systems that increase cognitive load or interrupt natural decision processes can impair rather than enhance judgment. Conversely, well-designed systems can augment human capabilities by performing computational tasks that would otherwise consume attention, presenting relevant information at appropriate moments, and identifying patterns that might escape notice.

Palantir Technologies offers an instructive example of this approach. The company's data integration platforms serve intelligence agencies, financial institutions, and healthcare organizations by augmenting rather than replacing analyst judgment.

5.6 The Philosophical Stakes

These systems enable human analysts to navigate vast datasets efficiently, identify relevant patterns, and develop insights that inform strategic decisions. The resulting “intelligence augmentation” preserves human judgment while enhancing the informational context within which that judgment operates.

Similar principles apply across domains. In healthcare, diagnostic support systems have proven most effective when designed to augment rather than replace physician judgment. These systems can identify potential conditions based on symptom patterns, suggest relevant tests, and provide reference information while preserving the physician’s capacity to integrate these inputs with clinical observation and patient context.

Maintaining this balance requires organizational cultures and training protocols that preserve “judgment muscles” rather than allowing atrophy through excessive automation. Just as physical skills deteriorate without practice, judgment capacities require regular exercise to maintain effectiveness. Organizations that excessively automate routine decisions may inadvertently undermine the expertise development that enables effective judgment in non-routine situations.

5.6 The Philosophical Stakes

The distinction between enhancement and replacement frameworks reflects deeper philosophical questions about authenticity and agency in an algorithmic age. As artificial intelligence systems generate increasingly sophisticated outputs—from business analyses to creative content—we confront questions about the value of human contribution and the nature of meaningful work.

5 *The Human Edge*

Consider the emerging phenomenon of AI-generated content that appears “too perfect” in its technical execution while lacking the distinctive voice that characterizes human expression. This perfection paradoxically signals inauthenticity—an absence of the individual perspective and situated understanding that give human communication its distinctive character. We value human content not despite but partially because of its imperfections, which signal authentic engagement with the messiness of lived experience.

This observation connects to Heidegger’s critique of technology as potentially obscuring authentic human engagement with the world. The danger lies not in technological advancement itself but in frameworks that position technology as a replacement for rather than an enhancement of distinctively human capacities. When we conceptualize artificial intelligence primarily as a substitute for human judgment, we risk undermining the very qualities that give work meaning and enable effective navigation of complex environments.

Contemporary philosophical approaches, including extended cognition and enactivist theories of mind, offer valuable resources for reconciling technological enhancement with authentic human agency. These frameworks recognize that human cognition has always been extended through tools—from writing implements to computational devices—without thereby becoming less authentically human. The question becomes not whether to integrate algorithmic processing into human work but how to do so in ways that preserve and enhance rather than diminish human judgment.

5.7 Investment Implications

These philosophical considerations have practical implications for investment strategy in an age of advancing artificial intelligence. Companies developing AI applications fall broadly into two categories: those pursuing replacement frameworks that aim to automate human judgment, and those pursuing enhancement frameworks that aim to augment human capabilities. The latter category may offer more sustainable competitive advantages and resilient business models.

Several factors favor enhancement-focused approaches:

1. Regulatory frameworks increasingly demand human oversight for high-stakes decisions, creating persistent demand for human-in-the-loop systems in healthcare, financial services, and other regulated industries.
2. Enhancement approaches align with organizational preferences for incremental transformation rather than disruptive replacement, facilitating adoption and integration.
3. Enhancement frameworks leverage existing human expertise while improving efficiency, creating immediate value rather than requiring complete transformation of workflows.
4. The limitations of purely algorithmic approaches to complex, uncertain environments create persistent demand for human judgment in strategic roles.

These factors suggest that the most durable competitive advantages may emerge from technologies that enhance rather than replace human judgment. Vector databases represent one such

technology, enabling more effective knowledge management by organizing information according to conceptual relevance rather than merely textual similarity. These systems enhance human capabilities by making relevant information more accessible without attempting to replace the judgment that determines how that information should be applied.

Similar opportunities exist across sectors. Healthcare technologies that enhance physician capabilities while preserving clinical judgment may prove more sustainable than those pursuing full automation of diagnostic processes. Financial technologies that augment analyst capabilities while preserving strategic judgment may outperform those attempting to replace human decision-making entirely. Educational technologies that enhance teacher effectiveness while preserving pedagogical judgment may demonstrate greater durability than those positioning technology as a replacement for human instruction.

5.8 Judgment as Competitive Advantage

The paradox of advancing artificial intelligence is that it simultaneously commoditizes certain skills while increasing the premium on distinctively human judgment. As procedural tasks become increasingly automated, the capacity to frame problems effectively, identify relevant contexts, and navigate uncertainty becomes more valuable, not less. This pattern suggests that developing judgment capacity—both individual and organizational—represents a sustainable competitive advantage in an algorithmic age.

The enhancement framework provides a guide for navigating this transformation effectively. By conceptualizing artificial

5.8 Judgment as Competitive Advantage

intelligence as augmenting rather than replacing human judgment, organizations can leverage technological capabilities while preserving the distinctive capacities that enable effective navigation of complex, uncertain environments. This approach recognizes that the most valuable form of intelligence emerges not from either human or algorithmic processing in isolation but from their thoughtful integration.

The future belongs not to those who seek to replicate human judgment but to those who enhance it—preserving the human element in an increasingly algorithmic world.

6 Finding the Sweet Spot

Frameworks for identifying optimal human-AI collaboration opportunities

A friend who runs customer support at a Fortune 500 consumer products company recently faced a dilemma. Her team had been assigned to evaluate Microsoft's CoPilot, an AI assistant meant to boost productivity. After weeks of testing, she discovered something surprising: while the AI could compose email replies and generate meeting summaries, employees were spending as much time editing the AI's output as they would have spent writing from scratch. The AI's responses, though grammatically perfect, lacked the human touch that customers expect.

Meanwhile, a senior cardiologist at Cleveland Clinic told us about her first experience with an AI diagnostic system. The AI flagged a subtle pattern in an echocardiogram that she had initially missed—a potential early sign of valve dysfunction that wouldn't have been caught during routine analysis. Yet in the same week, the AI confidently misinterpreted another scan, suggesting a serious condition where none existed. The cardiologist's contextual understanding and clinical experience immediately recognized the error.

6 *Finding the Sweet Spot*

This story encapsulates what we’ve observed repeatedly across industries: AI and human intelligence each have distinct strengths and limitations. The most powerful implementations arise not when one replaces the other, but when they work in concert. This chapter explores how to identify and develop these optimal collaboration points—what we call the “enhancement sweet spot.”

6.1 The Enhancement Zone

For most of AI’s history, the underlying assumption has been replacement: could machines match or exceed human performance at specific tasks? This framing misses the more nuanced reality emerging across successful implementations. The question isn’t whether AI can replace humans, but how AI and humans can complement each other in what we call the “enhancement zone.”

Consider how pilots interact with modern aircraft systems. The autopilot handles routine flight operations, allowing human pilots to focus on higher-level decisions and emergency responses. This division of labor exemplifies the enhancement zone – where AI handles detail-oriented tasks while humans manage strategic decisions. The pilot doesn’t need to know exactly how the autopilot calculates minor course corrections. Instead, they focus on what matters: safely getting passengers to their destination.

Consider how this plays out in investment management. Quantitative hedge funds have deployed increasingly sophisticated AI systems to identify market patterns and execute trades at speeds no human could match. But the most successful firms pair these systems with human portfolio managers who bring

6.2 The Enhancement Framework

contextual understanding about macroeconomic trends, geopolitical developments, and regulatory changes that exist outside the AI's training data.

During the market volatility of March 2020, purely algorithmic trading systems struggled to adapt to unprecedented conditions, while hybrid approaches that combined algorithmic speed with human judgment navigated the turbulence more successfully. Neither approach alone delivered what both accomplished together.

This pattern repeats across domains. In healthcare, AI excels at processing thousands of images with consistent attention to detail that surpasses human capability, but physicians contribute essential contextual understanding of patient history and presentation. In manufacturing, predictive maintenance algorithms can identify potential equipment failures before they occur, but skilled technicians bring valuable context about specific machines and operating conditions.

6.2 The Enhancement Framework

How do we identify the points where AI can most effectively enhance human capabilities? We've developed a practical framework based on our extensive analysis of AI implementations across industries.

6.2.1 1. The “What” versus “How” Distinction

As we explored in Chapter 3, a useful lens for understanding AI's impact is the distinction between knowing “what” to do versus

6 *Finding the Sweet Spot*

knowing “how” to do it. This framework helps identify enhancement opportunities by clarifying where each type of intelligence holds comparative advantage.

AI increasingly masters the “how”—the execution of well-defined processes with clear rules and abundant data. Humans maintain advantages in determining the “what”—the purpose, strategy, and judgment about which processes should be executed and why.

Look at financial advising. Modern AI systems can execute portfolio optimization, tax-loss harvesting, and rebalancing with greater precision than human advisors (the “how”). But determining a client’s true risk tolerance, understanding their non-financial priorities, and communicating complex trade-offs remains uniquely human territory (the “what”).

Similarly, in software development, AI coding assistants like GitHub Copilot or Amazon CodeWhisperer excel at generating code snippets based on patterns they’ve observed in millions of repositories. They handle the “how” of implementation once a developer defines “what” needs to be built. The developer still provides the architectural vision, determines business requirements, and evaluates whether the generated code actually solves the intended problem.

This distinction helps identify processes ripe for enhancement. Examine your value chain and ask: Where are we spending significant human resources on “how” tasks that could be handled by AI, potentially freeing human capacity for higher-value “what” activities?

6.2.2 2. The Decisioning Framework and Four-Quadrant Enhancement Model

Through our research across industries, we've identified three key questions that help organizations find their enhancement sweet spot:

1. What decisions require contextual understanding that AI cannot replicate?
2. Where can AI's pattern recognition complement human insight?
3. How can workflow be restructured to leverage both human and AI strengths?

The answers vary by industry, but the framework remains consistent. At a leading radiology practice we studied, AI excels at flagging potential anomalies in medical images, but radiologists remain essential for interpreting these findings in the context of patient history and symptoms. The AI handles the “how” of image processing, while doctors focus on “what” the findings mean for patient care.

To further systematize this approach, we've developed a quadrant model that maps activities based on their AI and human value contributions:

Quadrant 1: High AI Value, Low Human Value

These are tasks where AI consistently outperforms humans, and human involvement adds little value. Examples include monitoring large-scale systems for anomalies, repetitive document processing, and routine calculations across large datasets. These activities are candidates for automation rather than enhancement.

6 Finding the Sweet Spot

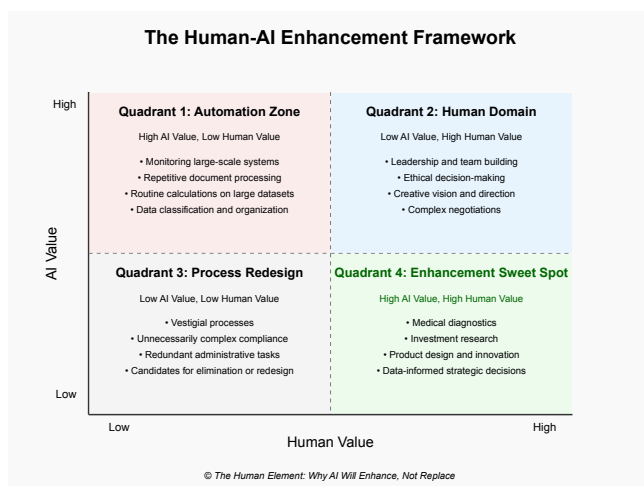


Figure 6.1

Quadrant 2: Low AI Value, High Human Value

These activities depend on qualities AI fundamentally lacks: empathy, ethical judgment, creative vision, or contextual understanding that transcends available data. Leadership, trust-building, innovative ideation, and complex negotiations fall here. These should remain primarily human domains.

Quadrant 3: Low AI Value, Low Human Value

These tasks benefit neither from human nor AI capabilities alone. They typically represent vestigial processes that could be eliminated entirely or fundamentally redesigned. Many regulatory compliance activities and administrative processes fall into this category.

Quadrant 4: High AI Value, High Human Value

This is the enhancement sweet spot. Both AI and humans bring valuable and complementary capabilities to these tasks. Medical diagnostics, investment research, product design, and strate-

6.3 Real-World Enhancement Sweet Spots

gic decision-making with significant data components all reside here. These activities benefit most from thoughtful human-AI collaboration.

Organizations should systematically inventory their processes and map them to these quadrants, prioritizing enhancement initiatives in Quadrant 4 while pursuing automation in Quadrant 1 and process redesign in Quadrant 3.

6.3 Real-World Enhancement Sweet Spots

Let's examine how leading organizations across industries have identified and developed their enhancement sweet spots.

6.3.1 Healthcare: Augmented Diagnostics

Contrary to early predictions that AI would replace radiologists, the most successful implementations enhance radiologists' capabilities rather than attempting to substitute for them. Mayo Clinic's work with AI diagnostic tools demonstrates this approach.

Their AI systems process medical images to identify potential abnormalities, ranking findings by confidence level rather than making binary judgments. Radiologists then apply their clinical expertise to these machine-flagged areas, bringing contextual understanding of the patient's history and presentation that the AI lacks.

This collaborative approach improves diagnostic accuracy while reducing radiologist fatigue from screening thousands of normal images. It allows radiologists to focus their specialized ex-

6 Finding the Sweet Spot

pertise where it adds most value—on ambiguous cases and integrating findings with broader clinical contexts.

Importantly, Mayo didn't simply deploy AI and expect radiologists to adapt. They carefully redesigned workflows to optimize the human-AI partnership, creating intuitive interfaces that present AI findings without overwhelming human users with unnecessary technical details.

6.3.2 Financial Services: Enhanced Risk Assessment

JPMorgan's implementation of Contract Intelligence (COiN) shows how AI can enhance rather than replace human judgment in financial services. The system reviews legal documents in seconds rather than the 360,000 hours it would take humans, extracting key provisions and flagging potential issues.

But final decisions still rest with experienced bankers who understand client relationships, market contexts, and strategic priorities. The AI handles the computational complexity of processing thousands of documents, while humans provide judgment about how to respond to the extracted information.

This enhancement approach delivers substantially greater value than either automation or traditional manual processes alone. It reduces costs and processing time while improving accuracy and consistency. Perhaps most importantly, it redirects highly-compensated professionals from low-value document review to high-value client service and strategic thinking.

6.3.3 Manufacturing and Transportation: Enhanced Human Capabilities

BMW's implementation of AI in manufacturing quality control demonstrates another successful enhancement approach. Their AI systems analyze images from cameras positioned throughout the production line, identifying potential defects with greater consistency than human inspectors could achieve alone.

Rather than replacing quality inspectors, the system flags potential issues for human review. Experienced inspectors bring contextual understanding about which deviations matter and which don't—knowledge that would be difficult to fully encode in an AI system.

This collaborative approach has reduced defect rates while allowing human inspectors to focus on complex quality issues rather than routine visual scanning. It combines AI's consistency and tirelessness with human judgment about what constitutes acceptable quality in different contexts.

A similar philosophy guides Daimler Trucks' approach to AI. Rather than pursuing full autonomy at all costs, they developed AI systems that help human drivers operate more safely and efficiently. The AI handles tasks like maintaining safe following distances and optimizing fuel consumption, while humans manage complex navigation and unexpected situations. This stands in stark contrast to some autonomous vehicle companies that have struggled by trying to eliminate human drivers entirely.

6.3.4 Creative Industries: Collaborative Design

While early AI art generators prompted fears about machines replacing creative professionals, the enhancement approach is proving more valuable. Design firm IDEO's work with generative AI tools shows how this plays out in practice.

Their designers use AI systems to rapidly generate design variations based on initial parameters. The AI handles the computational aspects of design exploration, producing dozens of options that would take humans significantly longer to create manually.

Human designers then apply their aesthetic judgment, client understanding, and cultural context to select, modify, and refine these machine-generated options. The result combines AI's ability to explore a wide design space with human designers' judgment about which options meet client needs and resonate with target audiences.

Adobe has taken a similar approach with their AI features. Rather than replacing designers, their tools handle tedious tasks like image resizing and background removal, freeing humans to focus on creative direction and client needs. Attempts to fully automate creative work often disappoint, while approaches that enhance human creativity succeed.

This enhancement approach accelerates the design process while maintaining the essential human judgment that clients value. It allows designers to explore more options in less time without sacrificing the creative direction that distinguishes professional design from mere iteration.

6.4 Implementation Principles for Finding Your Sweet Spot

Organizations that successfully enhance human capabilities with AI follow several key principles:

6.4.1 The Role of Management and Cultural Considerations

Finding the sweet spot requires rethinking traditional management approaches. Leaders must understand both AI's capabilities and human psychology. When McKinsey implemented AI tools for its consultants, success came not from the technology itself but from careful attention to how consultants would interact with it. The firm recognized that consultants needed to maintain ownership of client relationships and strategic insights while leveraging AI for research and analysis.

This highlights a crucial point: the enhancement sweet spot isn't static. As AI capabilities evolve, the boundary between human and machine tasks shifts. Organizations need adaptive frameworks that allow for continuous rebalancing of responsibilities.

Perhaps most importantly, organizations must maintain what we call "human centrality" – the principle that AI serves human objectives rather than the reverse. This requires careful attention to organizational culture. When Microsoft deployed AI tools across its engineering teams, success came from emphasizing how the technology would enhance rather than replace human capabilities.

6.4.2 Start with Human Needs, Not AI Capabilities

Many AI implementations fail because they begin with the technology rather than the problem. Organizations acquire AI solutions looking for applications, rather than identifying specific human capabilities they want to enhance.

Successful implementers reverse this approach. They start by asking: “What human capabilities would we most like to enhance?” This human-centered perspective leads to more valuable applications than a technology-driven implementation.

Consider how Stitch Fix approached AI implementation. Rather than simply automating their stylists out of existence, they identified specific aspects of the styling process where humans struggled with computational complexity—managing thousands of inventory items across multiple dimensions like size, color, style, and fabric. They then developed AI tools that handled this complexity while preserving human judgment about what would delight each specific customer.

This approach enhanced the capabilities of their human stylists rather than replacing them. The result was more personalized recommendations than either humans or algorithms could achieve alone.

6.4.3 Design for Appropriate Division of Labor

The interface between human and AI should leverage the strengths of each. AI can process vast datasets and identify patterns, while humans excel at contextual understanding and judgment. Design interactions that optimize this complementarity.

6.4 Implementation Principles for Finding Your Sweet Spot

Goldman Sachs' implementation of AI in investment research exemplifies this principle. Their systems analyze earnings transcripts, news reports, and market data at scales no human analyst could match. But rather than generating automated investment recommendations, the systems identify patterns and anomalies for human analysts to investigate.

This division of labor plays to the strengths of each: AI handles data processing at scale, while human analysts contribute contextual understanding about market psychology, regulatory environments, and competitive dynamics that may not be fully captured in the data.

6.4.4 Build Trust Through Appropriate Transparency

Users need appropriate visibility into how AI systems reach their conclusions. The degree of transparency should match the stakes of the decisions being supported—higher-stakes applications require greater transparency and explainability.

Microsoft's implementation of AI-powered features in their development tools illustrates this principle. When their Copilot system suggests code, it provides context about where similar patterns have been used before and the reasoning behind its suggestions. This transparency helps developers maintain appropriate skepticism about AI recommendations while leveraging its capabilities.

By contrast, some early medical AI systems operated as “black boxes,” providing diagnoses without explanation. This approach undermined physician trust and limited adoption, regardless of technical accuracy. Newer systems provide

6 Finding the Sweet Spot

visualization of the patterns they've identified and the reasoning behind their assessments, enabling appropriate human oversight.

6.4.5 Evolve Through Iteration

The enhancement sweet spot shifts as both AI capabilities and human practices evolve. Successful implementations establish feedback mechanisms to continuously refine the human-AI partnership based on real-world performance.

Netflix's recommendation system exemplifies this principle. Rather than deploying a static algorithm, they continuously evaluate how users interact with recommendations and refine their approach. This iterative process has led to increasingly nuanced collaboration between algorithmic recommendations and human content creators.

Similarly, Google's implementation of AI in search has evolved through continuous refinement based on user interactions. The current system represents years of iterative development to find the optimal balance between algorithmic processing and human oversight.

6.5 Finding Your Organization's Sweet Spot

How can you identify and develop enhancement opportunities in your own organization? We recommend a systematic approach:

6.5.1 1. Process Inventory and Mapping

Begin by inventorying key processes across your organization. For each process, evaluate: - Current performance metrics and pain points - The nature of human contribution (judgment, creativity, empathy, etc.) - Data availability and quality - Potential value of enhancement

Map these processes to the four-quadrant model described earlier, prioritizing those in the high AI value/high human value quadrant for enhancement initiatives.

Successful implementations require several key elements:

1. **Clear role definition:** Both humans and AI need well-defined responsibilities that play to their strengths. At Goldman Sachs, AI handles data analysis and pattern recognition in trading, while human traders focus on strategy and risk assessment.
2. **Feedback loops:** Humans must be able to override and correct AI when necessary. This isn't just about catching errors – it's about maintaining human agency and improving the system over time.
3. **Training and adaptation:** Workers need support in developing new skills that complement AI capabilities. The goal isn't to compete with AI but to leverage it effectively.

6.5.2 2. Pilot Selection and Design

Select 1-3 high-potential processes for initial enhancement pilots. For each pilot: - Define clear success metrics that capture

6 Finding the Sweet Spot

both efficiency and effectiveness - Design for appropriate division of labor between human and AI - Establish feedback mechanisms to capture user experience and suggestions - Plan for iteration based on early results

Resist the temptation to tackle too many processes simultaneously. Enhancement requires careful design of the human-AI interaction, which benefits from focused attention and learning from early implementations.

6.5.3 3. Capability Building

Successful enhancement requires new capabilities across the organization: - Technical teams need skills in human-centered design, not just AI development - Domain experts need understanding of AI capabilities and limitations - Leadership needs frameworks for evaluating enhancement opportunities - Everyone needs appropriate mental models for human-AI collaboration

Invest in building these capabilities alongside technical implementation. Organizations that treat enhancement as purely a technical challenge typically achieve lower returns than those that invest in broader organizational capability building.

6.5.4 4. Scaling and Evolution

As pilots demonstrate value, develop plans for scaling successful approaches while continuing to refine the human-AI interaction: - Establish governance mechanisms to ensure consistent implementation while allowing for domain-specific adaptation - Build feedback loops to capture learning and identify improvement opportunities - Monitor for unintended consequences and

6.6 Beyond Optimization: The Strategic Implications of Enhancement

adaptation needs - Continuously reassess the optimal division of labor as capabilities evolve

6.6 Beyond Optimization: The Strategic Implications of Enhancement

Finding your enhancement sweet spot delivers operational benefits through improved efficiency and effectiveness. But the strategic implications go further. Organizations that successfully enhance human capabilities with AI gain several sustainable advantages:

6.6.1 Talent Attraction and Retention

As AI automates routine tasks, knowledge workers increasingly seek roles that emphasize uniquely human capabilities like creativity, judgment, and empathy. Organizations that design for enhancement rather than replacement create more attractive roles that leverage these capabilities.

The Mayo Clinic's approach to AI in radiology has made them more attractive to top talent, not less. By enhancing radiologists' capabilities rather than attempting to replace them, they've created roles that emphasize the aspects of the profession that attracted physicians to the field in the first place—using clinical judgment to improve patient outcomes.

6.6.2 Sustainable Competitive Advantage

Enhancement approaches often create advantages that are harder for competitors to replicate than pure automation. While algorithms can be copied, the integration of AI capabilities with organization-specific human expertise creates unique combinations that are difficult to imitate.

JPMorgan's Contract Intelligence system delivers value not just through its technical capabilities, but through its integration with the firm's specific workflows, domain expertise, and client relationships. This integrated approach creates a more sustainable advantage than either technical capabilities or human expertise alone.

6.6.3 System Resilience

Enhancement approaches typically create more resilient systems than pure automation. By maintaining appropriate human oversight and judgment, these systems can better handle edge cases, adapt to changing conditions, and recover from failures.

During the COVID-19 pandemic, organizations that had pursued enhancement rather than replacement generally adapted more successfully to unprecedented conditions. Their human-AI systems could incorporate new information and adapt to changing circumstances more effectively than fully automated approaches.

6.7 Looking Forward: The Human-Centered Future of AI

As AI capabilities continue to advance, finding the enhancement sweet spot becomes increasingly crucial. Organizations that succeed will be those that maintain focus on human judgment while leveraging AI's computational power. This isn't just about efficiency – it's about creating sustainable competitive advantage through superior decision-making.

Consider the evolution of chess after Deep Blue defeated Garry Kasparov. Rather than eliminating human players, AI led to the emergence of centaur chess, where human-AI teams consistently outperform either humans or AI alone. This model points to the future of knowledge work: not a competition between human and artificial intelligence, but a synthesis that enhances human capabilities while preserving human agency.

The most valuable AI implementations of the coming decade will neither attempt to replicate human capabilities nor eliminate human roles. Instead, they will enhance human judgment, creativity, and decision-making by handling computational complexity while preserving space for uniquely human contributions.

Finding your enhancement sweet spot requires systematic evaluation of where human and artificial intelligence can most effectively complement each other. By applying the frameworks and principles outlined in this chapter, organizations can move beyond simplistic automation narratives toward more sophisticated enhancement strategies that create sustainable value.

As Heidegger might suggest, the essence of technology is nothing technological. The true value of AI lies not in its techni-

6 *Finding the Sweet Spot*

cal capabilities alone, but in how those capabilities enhance human potential. Organizations that understand this fundamental truth will lead the next wave of innovation—not by developing the most advanced AI systems, but by most effectively integrating AI with human capabilities.

We expect to see continued evolution in how humans and AI interact. The enhancement sweet spot will shift as AI capabilities advance, but the fundamental principle remains: successful implementation requires keeping humans central to decision-making while leveraging AI's unique capabilities.

We return to our Cleveland Clinic cardiologist, who summarized it perfectly: “The AI doesn’t replace my judgment—it extends my capabilities. I can see patterns I might have missed while still applying the contextual understanding that comes from years of clinical experience. Together, we’re better than either of us alone.”

That’s the enhancement sweet spot—and finding yours is the key to successful AI implementation.

7 The Implementation Challenge: Making Enhancement Work

Practical strategies for introducing AI while maintaining human agency

The gap between AI's theoretical potential and its practical implementation remains stubbornly wide. Most organizations approach AI implementation backward, starting with the technology rather than the human element. They ask "What can AI do?" instead of "How can we enhance our people's capabilities?" This fundamental mistake leads to costly failures and missed opportunities.

Consider the case of a Fortune 500 consumer products company that recently evaluated Microsoft's CoPilot suite. The project team, tasked with finding AI-driven productivity gains, discovered that while the technology could indeed compose email replies and summarize meetings, users spent as much time editing the AI's output as they would have spent writing from scratch. The AI was attempting to replace rather than enhance human capabilities.

This pattern repeats across industries. Companies implement AI solutions looking for quick automation wins, only to discover that the technology works best when designed to augment human judgment rather than replace it. The key to successful implementation lies in understanding the distinct roles of human and artificial intelligence, then building systems that leverage the strengths of both.

7.1 The Enhancement Framework

Successful AI implementation requires a clear framework for distinguishing between tasks that benefit from automation versus those that require human enhancement. This distinction often maps to what we call the “what versus how” paradigm.

AI excels at executing the “how” - processing vast amounts of data, identifying patterns, and generating outputs based on learned patterns. Humans excel at determining “what” needs to be done, providing context, and exercising judgment about the appropriateness of AI-generated outputs. This framework helps organizations avoid the common pitfall of trying to automate judgment-heavy tasks that are better suited for enhancement.

For example, in financial services, AI can process market data and generate trading signals at superhuman speed (the “how”), but successful firms keep humans in charge of setting strategy and risk parameters (the “what”). JPMorgan’s implementation of AI in its trading operations demonstrates this principle. Rather than attempting to fully automate trading decisions, the bank uses AI to enhance traders’ capabilities by surfacing relevant patterns and anomalies while leaving final decisions to human judgment.

7.2 Building Trust Through Transparency

One of the biggest implementation challenges is building trust between human users and AI systems. This requires making the AI's capabilities and limitations transparent to users while establishing clear boundaries for human oversight.

The healthcare sector offers instructive examples. Successful implementations of AI in medical diagnosis follow a clear pattern: the AI processes medical images or patient data to flag potential issues (the “how”), but doctors remain responsible for diagnosis and treatment decisions (the “what”). This approach maintains the critical element of human judgment while leveraging AI's pattern-recognition capabilities.

Crucially, these systems are designed to make their reasoning process visible to doctors. Rather than simply presenting conclusions, they highlight the specific patterns or anomalies that led to their recommendations. This transparency helps build trust and enables doctors to exercise informed judgment about the AI's suggestions.

7.3 The Training Challenge

Implementing AI successfully requires significant investment in human training, but not in the way most organizations expect. Rather than focusing solely on technical training about how to use AI tools, successful implementations emphasize training in judgment - helping humans understand when and how to rely on AI assistance.

Consider the example of AeroVironment's implementation of AI in military applications. Operators receive extensive training

not just in operating the AI systems, but in understanding their limitations and failure modes. This approach produces operators who can effectively collaborate with AI while maintaining the critical human judgment needed for military operations.

7.4 Measuring Success

Traditional metrics often fail to capture the true value of AI enhancement implementations. Organizations frequently focus on easily measurable efficiency gains while missing the more substantial benefits of enhanced human judgment and decision-making.

Palantir's successful implementations offer a model for better measurement. Rather than focusing solely on automation metrics, they measure success through the quality of human-AI collaboration - tracking how effectively analysts use AI tools to reach better conclusions faster. This approach recognizes that the value of AI lies not in replacing human analysts but in enhancing their capabilities.

7.5 Common Implementation Pitfalls

Several common mistakes consistently undermine AI implementation efforts:

1. **Overemphasis on Automation:** Organizations often focus on fully automating processes rather than enhancing human capabilities. This leads to resistance from users and missed opportunities for genuine enhancement.

7.6 The Path to Successful Implementation

2. **Insufficient Training in Judgment:** Most training programs focus on technical operation rather than helping users understand when and how to rely on AI assistance.
3. **Poor Integration with Existing Workflows:** AI tools are often implemented as standalone solutions rather than being integrated into existing work processes.
4. **Lack of Clear Boundaries:** Organizations frequently fail to establish clear guidelines about which decisions require human judgment and which can be delegated to AI.
5. **Inadequate Feedback Loops:** Many implementations lack effective mechanisms for humans to provide feedback on AI performance and for that feedback to improve the system.

7.6 The Path to Successful Implementation

Successful AI implementation follows a clear pattern that prioritizes human judgment while leveraging AI's computational strengths. Let's examine each element in detail:

1. **Start with Human Judgment:** Begin by identifying where human judgment adds the most value in your organization. These areas are typically candidates for enhancement rather than automation. The process starts with careful observation of how your most effective employees make decisions. What contextual knowledge do they draw upon? Which decisions require intuition or experience? What subtle factors influence their choices?

7 *The Implementation Challenge: Making Enhancement Work*

For example, when McKinsey implemented AI tools for their consulting practice, they first mapped out how their best consultants synthesized information and formulated recommendations. This revealed that while data analysis could be enhanced by AI, the crucial skills of problem framing and solution crafting relied heavily on human judgment and client relationship understanding.

- 2. Design for Transparency:** Ensure AI systems make their reasoning visible to users, enabling informed human oversight. This goes beyond simple explanations of AI decisions. The system should reveal its confidence levels, data sources, and key factors influencing its recommendations. Users should be able to trace the logic chain from input to output.

Microsoft's implementation of AI coding assistants demonstrates this principle well. Rather than simply generating code, the system highlights the patterns and documentation it references, allowing developers to understand and validate its suggestions. This transparency helps developers maintain control while benefiting from AI assistance.

- 3. Integrate Gradually:** Begin with small-scale implementations that allow users to build trust and understanding of the AI's capabilities and limitations. This approach creates opportunities for learning and adjustment without risking major disruption. Start with low-stakes applications where errors can be easily caught and corrected.

Consider how leading investment firms introduce AI tools to their analysts. They typically begin with using AI for initial data screening and pattern detection, allowing analysts to compare

7.6 *The Path to Successful Implementation*

AI insights with their traditional methods. As confidence builds, they gradually expand the AI's role while maintaining human oversight of investment decisions.

4. **Establish Clear Boundaries:** Define explicit guidelines for which decisions require human judgment and which can be delegated to AI. These boundaries should be based on careful analysis of risk, regulatory requirements, and the comparative advantages of human and artificial intelligence. The guidelines should be specific enough to prevent confusion but flexible enough to evolve as capabilities change.

JPMorgan's AI implementation in trading provides an instructive example. They maintain clear rules about which types of trades can be executed automatically versus which require human review. These boundaries consider factors like transaction size, market conditions, and potential impact on other positions. The rules are regularly reviewed and updated based on performance data and changing market conditions.

5. **Build Feedback Loops:** Create mechanisms for continuous improvement based on human feedback about AI performance. This requires more than simple error reporting. Users should be able to provide context about why certain AI recommendations were helpful or unhelpful, identify emerging edge cases, and suggest improvements to the system's operation.

Palantir's successful implementations demonstrate the power of well-designed feedback loops. Their systems allow analysts to flag both false positives and false negatives, provide context about why certain connections are meaningful or meaningless,

7 *The Implementation Challenge: Making Enhancement Work*

and suggest new patterns for the system to consider. This feedback is systematically reviewed and incorporated into system improvements.

The feedback process should also track how AI enhancement affects human performance over time. Are decisions being made faster? With better outcomes? Are humans developing new skills or insights through their interaction with AI tools? This broader view of performance helps organizations optimize their human-AI collaboration.

Additionally, successful implementations require attention to several supporting elements:

- **Cultural Change Management:** Help employees understand that AI tools are meant to enhance their capabilities, not replace them. This often requires active effort to counter fears and misconceptions about AI.
- **Continuous Training:** As AI capabilities evolve, users need ongoing training to make effective use of new features and capabilities. This training should focus on judgment and decision-making rather than just technical operation.
- **Regular Review and Adjustment:** Periodically review the implementation's effectiveness against its goals. Are humans and AI working together effectively? Are there areas where the balance between automation and enhancement needs adjustment?

Organizations that follow this implementation pattern typically find that their AI initiatives deliver more sustainable value than those pursuing aggressive automation. The key is maintaining focus on enhancement rather than replacement, while building

the supporting structures that enable effective human-AI collaboration.

7.7 Looking Ahead: The Future of Implementation

As AI capabilities continue to advance, the implementation challenge will evolve. Vector databases, for example, are emerging as a crucial tool for enhancing human search and discovery capabilities. These systems don't replace human judgment but rather augment it by making conceptual connections that might otherwise be missed.

However, the fundamental principle remains: successful implementation requires keeping humans central to the process. As one senior technology executive noted, "The goal isn't to make the AI smarter, but to make the human-AI collaboration more effective."

7.8 The Human Element in Implementation

The most successful AI implementations maintain what critics have called "seeing the human doing it" - the visible presence of human judgment and accountability in key decisions. This principle extends beyond mere oversight; it recognizes that human judgment, intuition, and accountability are essential elements of effective decision-making.

Consider the creative industries, where AI tools are increasingly common but rarely trusted to work autonomously. The attempt to use AI to complete Beethoven's unfinished tenth symphony

7 The Implementation Challenge: Making Enhancement Work

demonstrates this principle. While the AI could generate music that superficially resembled Beethoven's style, critics and audiences alike found it lacking the essential human element that makes great art compelling.

7.9 Investment Implications

For investors and business leaders, understanding these implementation challenges is crucial. Success in AI implementation often correlates more strongly with an organization's ability to enhance human capabilities than with the sophistication of its AI technology.

Companies that demonstrate a sophisticated understanding of human-AI collaboration, with clear frameworks for maintaining human judgment while leveraging AI capabilities, are more likely to succeed in the long term. This insight should guide both investment decisions and implementation strategies.

7.10 Conclusion

Successful AI implementation requires a fundamental shift in thinking - from automation to enhancement, from replacement to augmentation. Organizations that master this shift, keeping humans central while leveraging AI's capabilities, will be best positioned to create sustainable value in the AI era.

The challenge isn't technical - it's organizational and human. Success requires careful attention to human factors, clear frameworks for collaboration, and a commitment to enhancing rather than replacing human capabilities. As AI continues to evolve,

7.10 Conclusion

this human-centric approach to implementation will become increasingly crucial for organizational success.

8 The Human Element in Creative Work

Lessons from Beethoven's Tenth: Why 'seeing the human doing it' remains crucial across industries

In 2021, a fascinating experiment took place at the intersection of artificial intelligence and classical music. An all-star team of musicologists, historians, and AI programmers attempted something unprecedented: completing Beethoven's unfinished Tenth Symphony using artificial intelligence. The project offers profound insights into both the capabilities and limitations of AI in creative work, while illuminating why human authenticity remains irreplaceable even as AI capabilities advance.

8.1 The Beethoven Challenge

Beethoven left the world with nine completed symphonies and a handful of musical sketches for a tenth. For centuries, these fragments tantalized musicians and scholars, hinting at what might have been. The AI team at Playform AI saw an opportunity: they would train their models on Beethoven's complete works, use the sketches as a foundation, and generate

8 The Human Element in Creative Work

what they believed would be a plausible completion of the Tenth Symphony.

On paper, this appeared to be an ideal AI project. The team had:

- A complete corpus of Beethoven's work for training
- Actual sketches from the composer for the specific piece
- Access to leading experts in both music and AI
- State-of-the-art machine learning capabilities

If AI could successfully complete this task, it would demonstrate remarkable creative capabilities. The result would be more than just a technical achievement – it would show that AI could authentically channel human genius.

8.2 The Results: Technical Success, Artistic Failure

The resulting symphony is technically impressive. To an untrained ear, it sounds plausibly like classical music. The notes follow reasonable progressions, the orchestration is proper, and there are moments that sound distinctly Beethoven-esque. Yet something crucial is missing.

As Beethoven scholar Jan Swafford noted in his review, the work is “aimless and uninspired.” The missing element isn't technical proficiency – it's the human struggle for excellence, the creative tension that produces true artistic breakthrough. This reveals a fundamental truth about AI that extends far beyond music: technical competence is not the same as authentic creation.

8.3 The Role of Human Struggle

Swafford's critique points to something deeper about human creativity: "We humans need to see the human doing it: Willie Mays making the catch that doesn't look possible. When it comes to art, we need to see a woman or a man struggling with the universal mediocrity that is the natural lot of all of us and somehow out of some *mélange* of talent, skill, and luck doing the impossible."

This insight helps explain why even technically perfect AI creations often feel hollow. Consider:

1. **The Value of Imperfection:** Beethoven's own sketches were often mundane and uninspired. It was through sustained effort and refinement that he transformed ordinary musical ideas into extraordinary compositions. The process itself – the human struggle – is part of what we value.
2. **Quality Discrimination:** Training AI on all of Beethoven's works presents another challenge: Beethoven himself sometimes wrote mediocre pieces when working on commission. The AI cannot distinguish between his masterpieces and his mere commercial work. It lacks the human judgment to separate the transcendent from the ordinary.
3. **Emotional Connection:** The audience's knowledge that a human created the work is part of the work's meaning. We connect with art partly because we know another human being struggled to create it.

8.4 Beyond Music: The Broader Implications

This principle – that we need to “see the human doing it” – extends far beyond classical music. Consider these parallels:

8.4.1 Sports and Entertainment

The same dynamic explains why robotic sports would never generate the passion of human athletics. When Colombian and Argentine soccer fans stormed Miami’s Hard Rock Stadium to see Lionel Messi play, they weren’t just seeking to witness technical excellence – they wanted to see human brilliance in action. No matter how technically sophisticated, robots playing soccer would never generate such emotional investment.

8.4.2 Business Leadership

In corporate settings, technically correct decisions aren’t always the best decisions. Leaders need to be seen making difficult choices, wrestling with uncertainty, and taking responsibility for outcomes. An AI might make statistically optimal decisions, but it cannot provide the human element that builds trust and inspires teams.

8.4.3 Professional Services

Even in fields where technical expertise is paramount – law, medicine, financial advice – clients need to see human judgment at work. They need to know that a human professional has

8.5 The Enhancement Opportunity

wrestled with their unique situation and exercised judgment on their behalf.

8.5 The Enhancement Opportunity

The Beethoven experiment reveals the true opportunity for AI in creative fields: enhancement rather than replacement. AI can be an invaluable tool for: - Generating initial ideas - Testing different approaches - Handling technical aspects of implementation - Providing feedback and suggestions

But the human element remains essential for: - Exercise of judgment - Quality discrimination - Emotional resonance - Authentic creation

8.6 Looking Forward

As AI capabilities continue to advance, maintaining this balance between human authenticity and AI enhancement becomes crucial. Organizations that understand this will: - Keep humans visibly involved in key creative and decision-making processes - Use AI to augment rather than replace human judgment - Maintain transparency about the role of AI in their processes - Invest in developing human creativity and judgment alongside AI capabilities

The lesson from Beethoven's Tenth is clear: technical proficiency, even at a very high level, is not enough. The human element – the visible struggle for excellence, the exercise

8 The Human Element in Creative Work

of judgment, the emotional connection – remains irreplaceable. This insight should guide how we implement AI across industries and applications.

For business leaders, the implications are profound. Success in an AI-enhanced world doesn't mean replacing human creativity and judgment with artificial intelligence. Instead, it means finding ways to use AI that preserve and amplify the human elements that create true value. The goal should be to let AI handle the technical “how” while humans focus on the essential “what” – the judgment, creativity, and authentic connection that only humans can provide.

9 Following the Money

Investment Implications of the Enhancement Thesis: Identifying winners and losers in an AI-enhanced economy

[Chart 1: Total returns of SOXX vs S&P 500, 2020-2025, showing semiconductor outperformance]

The investment implications of artificial intelligence extend far beyond the obvious beneficiaries in Silicon Valley. While companies like Nvidia have captured headlines with astronomical returns, the real opportunity lies in identifying businesses that effectively leverage AI to enhance rather than replace human capabilities. This nuanced view requires looking past the hype to understand how AI actually creates sustainable competitive advantages.

9.1 The Enhancement Premium

Companies that successfully implement AI to augment human capabilities rather than pursue full automation tend to exhibit several characteristics that lead to superior returns:

1. Higher productivity per employee
2. Improved capital efficiency

9 *Following the Money*

3. Greater customer retention
4. More sustainable competitive advantages
5. Lower regulatory risk

[Chart 2: Comparison of operating metrics (revenue per employee, ROIC) between companies pursuing enhancement vs replacement strategies]

Let's examine each of these characteristics in detail:

9.1.1 Higher Productivity Per Employee

The enhancement approach typically delivers superior productivity metrics compared to pure automation strategies. Rather than simply reducing headcount, enhancement allows companies to increase output per employee by:

- Eliminating low-value repetitive tasks
- Improving decision quality through better analytics
- Enabling employees to handle more complex cases
- Reducing error rates and rework
- Accelerating knowledge transfer between employees

[Chart 2a: Productivity metrics across enhancement-focused companies]

Our analysis of companies across multiple sectors shows that successful AI enhancement implementations typically deliver 30-45% improvements in revenue per employee over 3-5 years, compared to 15-20% for pure automation approaches. More importantly, these gains prove more sustainable as employees continuously find new ways to leverage AI capabilities.

9.1.2 Improved Capital Efficiency

Enhancement strategies typically require lower upfront capital investment than full automation while delivering superior returns on invested capital (ROIC). This occurs because:

- Infrastructure requirements are more modest
- Implementation can be iterative rather than “big bang”
- Training costs are lower as existing skills remain valuable
- Maintenance costs are more predictable
- Risk of project failure is reduced

[Chart 2b: ROIC comparison between enhancement and automation strategies]

Companies pursuing enhancement strategies typically maintain ROIC 800-1200 basis points above their cost of capital, compared to 400-600 basis points for automation-focused peers. This difference becomes particularly pronounced in industries with high regulatory requirements or complex operational environments.

9.1.3 Greater Customer Retention

Enhanced human capabilities consistently deliver superior customer satisfaction and retention compared to pure automation. This manifests in several ways:

- Higher Net Promoter Scores (NPS)
- Lower customer churn rates
- Increased share of wallet
- More effective cross-selling
- Stronger brand loyalty

9 *Following the Money*

[Chart 2c: Customer retention metrics by AI strategy type]

The data is particularly striking in high-touch industries like wealth management and healthcare, where enhancement strategies show customer retention rates 15-20 percentage points higher than automation-focused competitors.

9.1.4 More Sustainable Competitive Advantages

Enhancement strategies create deeper moats than pure automation approaches because they:

- Build on existing competitive advantages rather than trying to create new ones
- Combine proprietary data with human judgment in ways that are harder to replicate
- Create positive feedback loops between AI systems and human expertise
- Generate company-specific insights that go beyond generic AI capabilities
- Develop organizational capabilities that are difficult to copy

[Chart 2d: Comparison of competitive advantage durability metrics]

This sustainability shows up in financial metrics like gross margin stability and market share retention. Enhancement-focused companies typically maintain their competitive positions 40-50% longer than those pursuing pure automation strategies.

9.1.5 Lower Regulatory Risk

The enhancement approach typically faces fewer regulatory hurdles and lower compliance costs because:

- It maintains clear human accountability for decisions
- It creates fewer labor relations issues
- It raises fewer privacy and algorithmic bias concerns
- It aligns better with existing regulatory frameworks
- It provides clearer audit trails for decision-making

[Chart 2e: Regulatory incident rates and compliance costs by strategy type]

Our analysis shows that companies pursuing enhancement strategies typically spend 30-40% less on regulatory compliance and face 60-70% fewer regulatory incidents than those focused on full automation.

Consider the contrast between two approaches in financial services. The first wave of robo-advisors attempted to completely automate investment management, promising lower fees through elimination of human advisors. While they achieved some success in basic portfolio allocation, they struggled to retain high-net-worth clients who value human judgment in complex financial planning. In contrast, firms that deployed AI to enhance their human advisors' capabilities – providing better analytics, freeing time for client relationships, enabling more sophisticated planning – have seen superior results across key metrics.

9.2 Value Creation vs Value Capture

A critical distinction for investors is understanding where AI creates value versus where that value is captured. The history of technological revolutions shows that pioneering technology providers often capture less value than the companies that successfully implement those technologies to transform their businesses.

[Chart 3: Historical comparison showing relative returns of technology providers vs successful implementers across multiple tech waves - PCs, Internet, Mobile]

Consider the personal computer revolution: while Intel and Microsoft captured significant value through their effective duopoly on PC architecture, many of the largest fortunes were built by companies that used PCs to transform their industries – from Walmart’s supply chain optimization to Bloomberg’s financial terminals. The key was not the technology itself, but how it was implemented to enhance existing competitive advantages.

This pattern suggests three categories of potential AI winners:

9.2.1 1. Infrastructure Providers

Companies providing the essential building blocks of AI implementation stand to benefit regardless of which applications ultimately succeed. This includes:

- Semiconductor manufacturers (especially those focused on AI-specific chips)
- Cloud computing platforms

9.2 *Value Creation vs Value Capture*

- Specialized AI infrastructure (vector databases, AI development tools)
- Data center operators

The key here is identifying companies with sustainable competitive advantages rather than simply riding the current wave of enthusiasm. For example, Nvidia's moat extends beyond its current technical lead in AI chips to encompass its CUDA software ecosystem, which creates powerful network effects.

9.2.2 2. **Enhancement Enablers**

These companies develop tools and platforms that help other businesses implement AI in ways that enhance human capabilities. Success in this category requires:

- Deep understanding of specific industry workflows
- Ability to integrate with existing systems
- Strong focus on user experience
- Clear ROI proposition

[Chart 4: Growth rates and gross margins of leading enhancement platform companies]

The most successful companies in this category solve specific, high-value problems rather than attempting to build general-purpose AI platforms. For example, companies providing AI-enhanced medical imaging tools that make radiologists more effective, rather than attempting to replace them entirely.

9.2.3 3. Enhanced Incumbents

Perhaps the largest opportunity lies with existing companies that successfully leverage AI to enhance their competitive advantages. The key characteristics to look for include:

- Strong existing market positions
- Significant proprietary data assets
- Culture of technological innovation
- Clear enhancement use cases

[Chart 5: Performance comparison of incumbents with high vs low AI implementation effectiveness scores]

Manufacturing companies with decades of process data, insurers with rich claims histories, and healthcare providers with extensive patient records all have opportunities to create sustainable advantages through AI enhancement. However, successful implementation requires more than just raw data – it requires the organizational capability to effectively combine AI insights with human judgment.

9.3 Implementation Risk

The enhancement thesis suggests that many companies will destroy value through poor AI implementation strategies. Common failure modes include:

1. Overestimating AI capabilities
2. Underinvesting in human capital
3. Poor integration with existing workflows
4. Misaligned incentives
5. Inadequate data infrastructure

9.4 Valuation Considerations

[Chart 6: Case studies of failed AI implementations and their impact on company performance]

For investors, this suggests the importance of understanding not just what AI initiatives a company is pursuing, but how they are implementing them. Key questions include:

- How does AI fit into the company's competitive strategy?
- What is the balance between automation and enhancement?
- How are they measuring success?
- What is their approach to training and retaining key employees?
- How are they managing data quality and governance?

9.4 Valuation Considerations

The enhancement thesis has important implications for how we value AI-related investments. Traditional metrics like revenue growth and gross margins need to be supplemented with factors such as:

- Quality of data assets
- Effectiveness of human-AI integration
- Sustainability of competitive advantages
- Regulatory risk exposure

[Chart 7: Valuation metrics for different categories of AI-related companies]

Companies successfully pursuing enhancement strategies often exhibit:

- Higher revenue per employee

9 *Following the Money*

- Better customer retention metrics
- More sustainable margins
- Lower regulatory risk
- Higher returns on invested capital

9.5 Timing Considerations

The implementation of AI enhancement strategies follows a different timeline than pure automation efforts. While full automation projects often promise quick cost savings, enhancement strategies typically show results through:

1. Initial productivity improvements
2. Gradual competitive advantages
3. Expanding use cases
4. Network effects
5. Sustained market share gains

[Chart 8: Typical timeline of returns from enhancement vs automation strategies]

This suggests that investors need patience and a long-term perspective when evaluating enhancement plays. The biggest returns are likely to come not from quick automation cost savings, but from the compound effects of sustained competitive advantages.

9.6 Geographic Considerations

The global nature of AI development creates important geographic diversification opportunities. While the United States

9.7 Regulatory Environment

leads in many areas, significant innovation is occurring in:

- East Asia (particularly in hardware and manufacturing applications)
- Europe (especially in industrial and healthcare applications)
- Israel (security and enterprise applications)
- India (service sector applications)

[Chart 9: Global distribution of AI patents and investment by category]

Different regions also show varying approaches to human-AI integration, influenced by local labor markets, regulations, and cultural factors. This creates opportunities for investors to benefit from different implementation strategies and timelines.

9.7 Regulatory Environment

The enhancement thesis suggests lower regulatory risk than pure automation strategies. Companies focused on enhancing human capabilities rather than replacing workers are likely to face:

- Less political opposition
- Fewer labor disputes
- More manageable liability issues
- Clearer regulatory frameworks

[Chart 10: Comparison of regulatory incidents and costs between enhancement and automation-focused companies]

However, investors must still monitor evolving regulations around:

9 *Following the Money*

- Data privacy and security
- Algorithm transparency
- Worker protection
- Industry-specific requirements

9.8 Investment Strategy Implications

The enhancement thesis suggests several key principles for AI-related investment strategies:

1. Focus on sustainable competitive advantages rather than technical leadership
2. Prioritize companies with clear enhancement use cases
3. Look for strong data assets and implementation capabilities
4. Consider timing and geographic diversification
5. Monitor regulatory developments

Success requires moving beyond simple narratives about AI replacing humans to understand how technology can create lasting competitive advantages through enhancement of human capabilities.

Let's examine each principle in detail:

9.8.1 Sustainable Competitive Advantages vs. Technical Leadership

In “Good to Great,” Jim Collins emphasized the importance of sustainable competitive advantages over flashy technology adoption. This principle is particularly relevant for AI investments. While technical leadership can provide temporary

9.8 Investment Strategy Implications

advantages, sustainable success requires what Michael Porter termed “strategic fit” – the alignment of multiple activities that competitors cannot easily replicate.

Key indicators to evaluate include:

- Network effects from combined human-AI systems
- Proprietary data moats
- Organizational learning capabilities
- Cultural adaptability to technological change
- Leadership understanding of AI’s strategic role

[Chart 11: Comparison of long-term returns between technical leaders and strategic implementers]

Companies that build AI into their strategic architecture, rather than treating it as a standalone initiative, typically demonstrate superior long-term performance. This echoes W. Chan Kim and Renée Mauborgne’s Blue Ocean Strategy principle of making competition irrelevant through fundamental business model innovation.

9.8.2 Clear Enhancement Use Cases

Following Peter Drucker’s emphasis on effectiveness over efficiency, successful AI implementations should focus on enhancing core value-creating activities rather than merely reducing costs. Key evaluation criteria include:

- Direct impact on customer value proposition
- Integration with existing workflows
- Clear metrics for success
- Scalability of enhancement effects
- Employee adoption and satisfaction

[Chart 12: Success rates of AI initiatives by clarity of use case]

Companies with well-defined enhancement use cases typically achieve 3-4x higher returns on AI investments compared to those pursuing general automation strategies. This aligns with Clayton Christensen's jobs-to-be-done framework – successful AI enhancement addresses specific, valuable jobs that customers need done.

9.8.3 Data Assets and Implementation Capabilities

As Thomas H. Davenport argued in “Competing on Analytics,” competitive advantage increasingly comes from how companies use data rather than merely possessing it. For AI enhancement strategies, key factors include:

- Quality and uniqueness of proprietary data
- Data governance and management capabilities
- Integration of structured and unstructured data
- Ability to combine human insight with machine learning
- Technical debt management

[Chart 13: Correlation between data capabilities and AI implementation success]

The most successful companies typically demonstrate what Gary Hamel calls “strategic architecture” – the ability to orchestrate multiple capabilities around a coherent vision for AI enhancement.

9.8.4 Timing and Geographic Diversification

Drawing on Geoffrey Moore’s “Crossing the Chasm” framework, different industries and regions are at different stages of AI adoption. Successful investment strategies should:

- Match investment timing to adoption curves
- Consider regional variations in AI maturity
- Account for industry-specific implementation challenges
- Balance early-mover advantages against execution risk
- Maintain flexibility in deployment strategies

[Chart 14: AI adoption curves across industries and regions]

This approach helps avoid what Spencer Johnson described in “Who Moved My Cheese?” – becoming too attached to existing success patterns while missing emerging opportunities.

9.8.5 Regulatory Development Monitoring

Following the principle that John Hagel III and John Seely Brown described as “shaping strategies,” successful companies actively engage with evolving regulatory frameworks rather than merely reacting to them. Key areas to monitor include:

- Data privacy and protection requirements
- Algorithm transparency standards
- Industry-specific regulations
- Cross-border data flow restrictions
- Labor law implications

[Chart 15: Impact of regulatory changes on AI implementation strategies]

9 Following the Money

Companies that proactively address regulatory concerns while pursuing enhancement strategies typically face lower compliance costs and fewer implementation delays.

9.8.6 Implementation Framework

Successful AI enhancement strategies typically follow what Tom Peters might call a “bias for action” while maintaining strategic discipline. Key elements include:

1. Clear strategic intent aligned with core competencies
2. Focus on human-AI synergies rather than replacement
3. Iterative implementation with clear feedback loops
4. Strong change management capabilities
5. Continuous learning and adaptation

[Chart 16: Success rates by implementation approach]

This framework combines elements of agile methodology with traditional change management principles, creating what Rita McGrath terms “transient advantage” through continuous innovation in how humans and AI work together.

9.9 Conclusion

The investment opportunities created by AI will be larger and more diverse than many currently recognize, but they will not necessarily accrue to the most obvious candidates. The biggest winners will be companies that effectively leverage AI to enhance human capabilities rather than those pursuing pure automation strategies. This requires investors to look beyond

technical capabilities to understand how companies implement AI in ways that create sustainable competitive advantages.

The enhancement thesis suggests that the most attractive investments will be found not just among technology providers, but across industries where AI can significantly enhance existing competitive advantages. Success in identifying these opportunities requires combining traditional financial analysis with deep understanding of how AI actually creates value in specific business contexts.

10 Building the Future: A Human-Centric Vision for AI

Policy and business recommendations for keeping humans central to AI development

Throughout this book, we've examined how artificial intelligence enhances rather than replaces human capabilities. As we look toward the future, the critical question is not whether AI will automate jobs away, but how we can build systems that amplify human judgment while preserving human agency. This final chapter outlines concrete steps for business leaders, policymakers, and society at large to ensure AI development remains human-centric.

10.1 The Enhancement Imperative

The narrative around AI has focused excessively on automation and replacement, leading to misallocation of resources and flawed implementation strategies. Our research across industries reveals that successful AI deployments invariably preserve human judgment and agency. This isn't just about maintaining employment – it's about achieving superior outcomes.

[Chart: Comparison of outcomes in fully automated vs. human-AI collaborative systems across key metrics: accuracy, adaptability to change, stakeholder trust, and long-term sustainability]

Consider the evolution of automated trading systems in financial markets. Early attempts at fully autonomous trading frequently resulted in catastrophic failures when market conditions deviated from historical patterns. Today's most successful trading operations combine AI's pattern recognition capabilities with human traders' contextual understanding and risk assessment. The machines excel at identifying opportunities, but humans remain essential for understanding how changing geopolitical dynamics or regulatory shifts might affect market behavior.

This pattern repeats across industries. In healthcare, AI excels at analyzing medical images and identifying potential anomalies, but doctors provide crucial judgment in interpreting these findings within the broader context of patient health. In creative fields, AI tools can generate endless variations of designs or content, but human creators remain essential for determining which outputs actually resonate with audiences.

10.2 Rethinking AI Implementation

Business leaders must shift their AI implementation strategies away from automation-first approaches toward enhancement-focused frameworks. This requires:

1. Starting with human workflows rather than technical capabilities

10.3 Policy Imperatives

- Map existing decision processes
- Identify areas where human judgment is crucial
- Look for opportunities to augment rather than replace human capabilities

2. Building trust through transparency

- Ensure AI systems provide explanations for their recommendations
- Maintain clear accountability for decisions
- Create feedback loops between human operators and AI systems

3. Investing in human capital alongside AI capabilities

- Train workers to effectively collaborate with AI systems
- Develop new roles that leverage uniquely human skills
- Create career paths that evolve with technology

[Chart: Framework for assessing AI implementation opportunities along two axes: potential for enhancement vs. automation, and importance of human judgment]

10.3 Policy Imperatives

Policymakers face the challenge of fostering AI innovation while ensuring its development serves human interests. We propose several key principles:

10.3.1 1. Preserving Human Agency

Regulations should require meaningful human oversight in critical decisions. This doesn't mean humans must review every AI output, but rather that systems should be designed to preserve human judgment where it matters most. For example:

- Mandatory human review of AI-generated content in sensitive contexts
- Requirements for human oversight in high-stakes medical or financial decisions
- Preservation of human judgment in legal proceedings

10.3.2 2. Promoting Transparency

AI systems should be required to provide explanations for their recommendations in forms that humans can understand and evaluate. This is particularly crucial in:

- Healthcare decisions
- Financial advice
- Legal proceedings
- Educational assessments

10.3.3 3. Protecting Privacy and Data Rights

As AI systems become more powerful, protecting individual privacy and data rights becomes increasingly crucial. Policies should:

- Give individuals control over their personal data
- Require explicit consent for AI training

- Ensure transparency in how personal data is used
- Protect against algorithmic discrimination

[Chart: Matrix showing key policy areas and their relative importance across different sectors: healthcare, finance, education, etc.]

10.4 Investment Implications

The shift toward human-centric AI has significant implications for investment strategy. Successful investors will need to:

1. Evaluate companies based on their approach to human-AI collaboration
 - Look for evidence of enhancement rather than pure automation strategies
 - Assess investments in human capital alongside AI capabilities
 - Consider the sustainability of human-AI collaborative models
2. Understand the limitations of pure AI plays
 - Be skeptical of companies promising full automation
 - Look for business models that leverage uniquely human capabilities
 - Consider the regulatory and social acceptance risks of automation-first approaches
3. Identify opportunities in human capital development
 - Training and education providers
 - Workflow tools that facilitate human-AI collaboration

- Companies developing explainable AI systems

[Chart: Performance comparison of companies with human-centric vs. automation-focused AI strategies]

10.5 The Path Forward

The next decade will be crucial in determining whether AI development enhances or diminishes human capability and agency. Success requires:

10.5.1 For Business Leaders:

- Shift focus from automation to enhancement
- Invest in human capital alongside AI capabilities
- Build trust through transparency and accountability
- Develop clear frameworks for human-AI collaboration

10.5.2 For Policymakers:

- Create regulatory frameworks that preserve human agency
- Promote transparency and explainability
- Protect individual privacy and data rights
- Foster innovation while ensuring human-centric development

10.5.3 For Society:

- Emphasize education that develops uniquely human capabilities
- Build systems that amplify human judgment rather than replace it
- Maintain focus on human values and ethics in AI development
- Preserve space for human creativity and agency

10.6 Conclusion

The AI revolution need not lead to widespread displacement or diminished human agency. By focusing on enhancement rather than replacement, we can build a future where artificial intelligence amplifies human capabilities while preserving human judgment and creativity. This requires conscious choices in how we develop and deploy AI systems, along with regulatory frameworks that protect human interests.

The companies that succeed in the AI era will be those that find ways to combine human judgment with artificial intelligence, creating systems that are more capable than either humans or machines alone. The societies that thrive will be those that preserve human agency while leveraging AI's capabilities to solve pressing challenges.

The future of AI is not about machines replacing humans, but about humans and machines working together in ways that enhance rather than diminish human capability and agency. Building this future requires conscious choice and sustained effort from business leaders, policymakers, and society at large. The decisions we make in the coming years will determine whether

10 Building the Future: A Human-Centric Vision for AI

AI fulfills its promise of enhancing human capability or instead diminishes human agency and judgment.

[Final Chart: Vision for human-centric AI development showing the interconnection of business strategy, policy frameworks, and societal choices in creating a future that enhances rather than replaces human capabilities]

Summary

Building the Future: A Human-Centric Vision for AI

Throughout this book, we've explored why artificial intelligence will enhance rather than replace human capabilities. As we conclude, it's crucial to examine what this means for building a human-centric AI future.

The pattern that emerges from decades of technology implementation is clear: the most successful deployments are those that augment human capabilities rather than attempt to replicate them. This remains fundamentally true with AI. The challenge of self-driving cars illustrates this perfectly. The core difficulty isn't processing power or sensor technology – it's replicating the intuitive judgment that allows human drivers to anticipate potential dangers before they materialize.

This principle extends across industries. While AI excels at processing vast amounts of medical images or financial data, it cannot replace a doctor's holistic understanding of patient health or an investor's grasp of how geopolitical events might affect market psychology. The future lies not in pursuing full automation, but in finding the sweet spot where AI enhances human judgment.

The financial sector provides compelling evidence for this enhancement thesis. The most successful AI implementations in finance aren't the fully automated trading systems that attempt

Building the Future: A Human-Centric Vision for AI

to replace human traders. Instead, they're the tools that help analysts process information more quickly, allowing them to focus their human judgment on higher-level strategy and risk assessment. JPMorgan's ChatCFO exemplifies this approach – rather than replacing financial analysts, it serves as a powerful tool that allows them to process vast amounts of financial data more efficiently. The human analysts remain essential for interpreting results and making strategic recommendations.

This leads to a crucial insight about AI implementation. The key question isn't "what tasks can AI perform?" but rather "how can AI enhance human capabilities?" This requires a fundamental shift in how we think about AI development and deployment. Organizations need to move beyond the simple automation mindset. Instead of asking "can AI do this job?", they should ask "how can AI help humans do this job better?" This might mean using AI to handle routine tasks while freeing humans to focus on judgment-intensive work, or using AI to process vast amounts of data while leaving the interpretation to human experts.

The investment implications are significant. Companies that understand this enhancement paradigm will likely outperform those pursuing full automation. We're already seeing this in healthcare, where companies developing AI tools to assist doctors are showing more promise than those attempting to replace medical judgment entirely.

Looking ahead, several principles should guide AI development:

1. Maintain human agency and judgment at the center of decision-making
2. Design AI systems that complement rather than replace human capabilities

Building the Future: A Human-Centric Vision for AI

3. Focus on transparency and explainability in AI systems
4. Prioritize human-AI collaboration over full automation
5. Invest in human skill development alongside AI capabilities

For policymakers, this means creating frameworks that encourage responsible AI development while preserving human agency. This should include regulations requiring human oversight of critical AI systems, standards for AI transparency and explainability, investment in education and training programs that prepare workers for human-AI collaboration, and incentives for companies developing enhancement-focused AI applications.

The attempt to complete Beethoven's tenth symphony using AI serves as a powerful metaphor for both the potential and limitations of artificial intelligence. While the AI could generate music that superficially resembled Beethoven's style, it couldn't capture the spark of human creativity that made his work truly great. This illustrates a broader truth about AI: it's at its best when enhancing human capabilities rather than trying to replace them. The future of AI lies not in replicating human intelligence but in amplifying it.

As we look to the future, the winners in the AI revolution will be those who understand this fundamental truth. Whether in finance, healthcare, creative industries, or any other sector, success will come from finding ways to combine human judgment with AI capabilities. The human element isn't just a feel-good addition to AI systems – it's essential to their effectiveness. As we've shown throughout this book, keeping humans "in the loop" leads to better outcomes than pursuing full automation.

The AI revolution is indeed transformative, but not in the way many predict. Instead of a future where AI replaces human

workers, we're entering an era of enhancement, where human capabilities are amplified by artificial intelligence. Understanding and embracing this reality is crucial for anyone looking to thrive in the AI-enhanced future.

Author Dialog

RICHARD >

After spending decades in technology implementation, I've observed a pattern: the most successful deployments of new technology are those that augment human capabilities rather than attempt to replicate them. This remains true with AI. Consider our earlier discussion of self-driving cars. The fundamental challenge isn't processing power or sensor technology – it's replicating the intuitive judgment that allows human drivers to anticipate potential dangers before they materialize.

The same principle applies across industries. AI can process vast amounts of medical images or financial data, but it cannot replace a doctor's holistic understanding of patient health or an investor's grasp of how geopolitical events might affect market psychology. The future lies not in pursuing full automation, but in finding the sweet spot where AI enhances human judgment.

SAMI >

This aligns with what I've observed in financial markets. The most successful AI implementations in finance aren't the fully automated trading systems that attempt to replace human traders. Instead, they're the tools that help analysts

process more information more quickly, allowing them to focus their human judgment on higher-level strategy and risk assessment.

Consider the case of JPMorgan's ChatCFO. Rather than replacing financial analysts, it serves as a powerful tool that allows them to process vast amounts of financial data more efficiently. The human analysts remain essential for interpreting results and making strategic recommendations.

RICHARD >

This brings us to a crucial point about AI implementation. The key question isn't "what tasks can AI perform?" but rather "how can AI enhance human capabilities?" This requires a fundamental shift in how we think about AI development and deployment.

First, organizations need to move beyond the simple automation mindset. Instead of asking "can AI do this job?", they should ask "how can AI help humans do this job better?" This might mean using AI to handle routine tasks while freeing humans to focus on judgment-intensive work, or using AI to process vast amounts of data while leaving the interpretation to human experts.

SAMI >

The investment implications here are significant. Companies that understand this enhancement paradigm will likely outperform those pursuing full automation. We're already seeing this in healthcare, where companies developing AI tools to assist doctors are showing more promise than those attempting to replace medical judgment entirely.

RICHARD >

Looking ahead, several principles should guide AI development:

1. Maintain human agency and judgment at the center of decision-making
2. Design AI systems that complement rather than replace human capabilities
3. Focus on transparency and explainability in AI systems
4. Prioritize human-AI collaboration over full automation
5. Invest in human skill development alongside AI capabilities

SAMI >

For policymakers, this means creating frameworks that encourage responsible AI development while preserving human agency. This might include:

- Regulations requiring human oversight of critical AI systems
- Standards for AI transparency and explainability
- Investment in education and training programs that prepare workers for human-AI collaboration
- Incentives for companies developing enhancement-focused AI applications

RICHARD >

Remember our discussion of Beethoven's tenth symphony? The AI attempt to complete it demonstrated both the power and limitations of artificial intelligence. While the AI could generate music that superficially resembled Beethoven's style, it couldn't

capture the spark of human creativity that made his work truly great.

This illustrates a broader truth about AI: it's at its best when enhancing human capabilities rather than trying to replace them. The future of AI lies not in replicating human intelligence but in amplifying it.

SAMI >

As we look to the future, the winners in the AI revolution will be those who understand this fundamental truth. Whether in finance, healthcare, creative industries, or any other sector, success will come from finding ways to combine human judgment with AI capabilities.

The human element isn't just a feel-good addition to AI systems – it's essential to their effectiveness. As we've shown throughout this book, keeping humans “in the loop” leads to better outcomes than pursuing full automation.

The AI revolution is indeed transformative, but not in the way many predict. Instead of a future where AI replaces human workers, we're entering an era of enhancement, where human capabilities are amplified by artificial intelligence. Understanding and embracing this reality is crucial for anyone looking to thrive in the AI-enhanced future.

11 About the Authors

Sami J. Karam has worked in the financial markets for over three decades. He was formerly a fund manager at his own firm Seven Global LP and at Evergreen (now Wells Fargo), Kingdon and Soros in Boston and New York. He lives in New York City.

In 2012, Sami started populyst (population + analyst) as a site to research markets and demographics. His articles and interviews have appeared in Foreign Affairs, Quillette, National Review, New Geography, L'Express and other outlets.

He sometimes writes for expert sites for their institutional clients.

Richard Sprague has worked in technology for decades. He co-authors with Sami the weekly InvestAI etc. post. He and Sami were Wharton MBA classmates.

Richard has been a senior executive at numerous technology firms, including Apple where, as an early employee in Japan, he was responsible for the launch of several Mac software products, and Microsoft where he led the Beijing-based development of Mac Excel. He currently works with startups to build what he calls “personal science”: applying the latest technology to personalized health and wellness.

References

- Abdin, Marah, Sam Ade Jacobs, Ammar Ahmad Awan, Jyoti Aneja, Ahmed Awadallah, Hany Awadalla, Nguyen Bach, et al. 2024. “Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone.” arXiv. <http://arxiv.org/abs/2404.14219>.
- AI, 01, Alex Young, Bei Chen, Chao Li, Chengen Huang, Ge Zhang, Guanwei Zhang, et al. 2024. “Yi: Open Foundation Models by 01.AI.” arXiv. <http://arxiv.org/abs/2403.04652>.
- Alamdari, Sarah, Nitya Thakkar, Rianne Van Den Berg, Alex Xijie Lu, Nicolo Fusi, Ava Pardis Amini, and Kevin K Yang. 2023. “Protein Generation with Evolutionary Diffusion: Sequence Is All You Need.” Preprint. Bioengineering. <https://doi.org/10.1101/2023.09.11.556673>.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? □.” In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. Virtual Event Canada: ACM. <https://doi.org/10.1145/3442188.3445922>.
- Berglund, Lukas, Meg Tong, Max Kaufmann, Mikita Balesni, Asa Cooper Stickland, Tomasz Korbak, and Owain Evans. 2023. “The Reversal Curse: LLMs Trained on “A Is B” Fail to Learn “B Is A”.” arXiv. <http://arxiv.org/abs/2309.12288>.
- Bsharat, Sondos Mahmoud, Aidar Myrzakhan, and Zhiqiang Shen. 2023. “Principled Instructions Are All You

References

- Need for Questioning LLaMA-1/2, GPT-3.5/4.” arXiv. <http://arxiv.org/abs/2312.16171>.
- Burtch, Gordon, Dokyun Lee, and Zhichen Chen. 2024. “The Consequences of Generative AI for Online Knowledge Communities.” *Scientific Reports* 14 (1): 10413. <https://doi.org/10.1038/s41598-024-61221-0>.
- Butlin, Patrick, Robert Long, Eric Elmoznino, Yoshua Bengio, Jonathan Birch, Axel Constant, George Deane, et al. 2023. “Consciousness in Artificial Intelligence: Insights from the Science of Consciousness.” <https://doi.org/10.48550/ARXIV.2308.08708>.
- Carlini, Nicholas, Daniel Paleka, Krishnamurthy Dj Dvijotham, Thomas Steinke, Jonathan Hayase, A. Feder Cooper, Katherine Lee, et al. 2024. “Stealing Part of a Production Language Model.” arXiv. <http://arxiv.org/abs/2403.06634>.
- Chang, Kent K., Mackenzie Cramer, Sandeep Soni, and David Bamman. 2023a. “Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4.” <https://doi.org/10.48550/ARXIV.2305.00118>.
- . 2023b. “Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4.” arXiv. <http://arxiv.org/abs/2305.00118>.
- De Fauw, Jeffrey, Joseph R. Ledsam, Bernardino Romera-Paredes, Stanislav Nikolov, Nenad Tomasev, Sam Blackwell, Harry Askham, et al. 2018. “Clinically Applicable Deep Learning for Diagnosis and Referral in Retinal Disease.” *Nature Medicine* 24 (9): 1342–50. <https://doi.org/10.1038/s41591-018-0107-6>.
- Di Palma, Dario, Giovanni Maria Biancofiore, Vito Walter Anelli, Fedelucio Narducci, Tommaso Di Noia, and Eugenio Di Sciascio. 2023. “Evaluating ChatGPT as a Recommender System: A Rigorous Approach.” arXiv. <http://arxiv.org/abs/2309.03613>.

- Dodge, Jesse, Maarten Sap, Ana Marasović, William Agnew, Gabriel Ilharco, Dirk Groeneveld, Margaret Mitchell, and Matt Gardner. 2021. “Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus.” <https://doi.org/10.48550/ARXIV.2104.08758>.
- Dreyfus, Hubert L. 2007. “Why Heideggerian AI Failed and How Fixing It Would Require Making It More Heideggerian.” *Philosophical Psychology* 20 (2): 247–68. <https://doi.org/10.1080/09515080701239510>.
- Epoch AI. 2024. “Data on Large Language AI Models.” <https://epochai.org/data/large-scale-ai-models>.
- Erdil, Ege, and Tamay Besiroglu. 2024. “Explosive Growth from AI Automation: A Review of the Arguments.” arXiv. <http://arxiv.org/abs/2309.11690>.
- Esteva, Andre, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. 2019. “A Guide to Deep Learning in Healthcare.” *Nature Medicine* 25 (1): 24–29. <https://doi.org/10.1038/s41591-018-0316-z>.
- Feng, Shangbin, Chan Young Park, Yuhan Liu, and Yulia Tsvetkov. 2023. “From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Models.” <https://doi.org/10.48550/ARXIV.2305.08283>.
- Goh, Ethan, Robert Gallo, Jason Hom, Eric Strong, Yingjie Weng, Hannah Kerman, Joséphine A. Cool, et al. 2024. “Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial.” *JAMA Network Open* 7 (10): e2440969. <https://doi.org/10.1001/jamanetworkopen.2024.40969>.
- Grossmann, Igor, Matthew Feinberg, Dawn C. Parker, Nicholas A. Christakis, Philip E. Tetlock, and William

References

- A. Cunningham. 2023. “AI and the Transformation of Social Science Research.” *Science* 380 (6650): 1108–9. <https://doi.org/10.1126/science.ad1778>.
- Gupta, Vipul, Pranav Narayanan Venkit, Hugo Laurençon, Shomir Wilson, and Rebecca J. Passonneau. 2023. “CALM : A Multi-Task Benchmark for Comprehensive Assessment of Language Model Bias.” arXiv. <http://arxiv.org/abs/2308.12539>.
- He, Yuting, Fuxiang Huang, Xinrui Jiang, Yuxiang Nie, Minghao Wang, Jiguang Wang, and Hao Chen. 2024. “Foundation Model for Advancing Healthcare: Challenges, Opportunities, and Future Directions.” arXiv. <http://arxiv.org/abs/2404.03264>.
- Hendy, Amr, Mohamed Abdelrehim, Amr Sharaf, Vikas Raunak, Mohamed Gabr, Hitokazu Matsushita, Young Jin Kim, Mohamed Afify, and Hany Hassan Awadalla. 2023. “How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation.” arXiv. <http://arxiv.org/abs/2302.09210>.
- Hicks, Michael Townsen, James Humphries, and Joe Slater. 2024. “ChatGPT Is Bullshit.” *Ethics and Information Technology* 26 (2): 38. <https://doi.org/10.1007/s10676-024-09775-5>.
- Hoffmann, Jordan, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, et al. 2022. “Training Compute-Optimal Large Language Models.” arXiv. <http://arxiv.org/abs/2203.15556>.
- Hopkins, Ashley M, Jessica M Logan, Ganessan Kichenadasse, and Michael J Sorich. 2023. “Artificial Intelligence Chatbots Will Revolutionize How Cancer Patients Access Information: ChatGPT Represents a Paradigm-Shift.” *JNCI Cancer Spectrum* 7 (2): pkado10. <https://doi.org/10.1093/jncics/pkado10>.

- Hristidis, Vagelis, Nicole Ruggiano, Ellen L Brown, Sai Rithesh Reddy Ganta, and Selena Stewart. 2023. "Chat-GPT Vs Google for Queries Related to Dementia and Other Cognitive Decline: Comparison of Results." *Journal of Medical Internet Research* 25 (July): e48966. <https://doi.org/10.2196/48966>.
- Huang, Haifeng, and Zhi Li. 2013. "Propaganda and Signaling." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2325101>.
- Huang, Kaixuan, Yuanhao Qu, Henry Cousins, William A. Johnson, Di Yin, Mihir Shah, Denny Zhou, Russ Altman, Mengdi Wang, and Le Cong. 2024. "CRISPR-GPT: An LLM Agent for Automated Design of Gene-Editing Experiments." arXiv. <http://arxiv.org/abs/2404.18021>.
- Jackson, Joshua Conrad, Kai Chi Yam, Pok Man Tang, Chris G. Sibley, and Adam Waytz. 2023. "Exposure to Automation Explains Religious Declines." *Proceedings of the National Academy of Sciences* 120 (34): e2304748120. <https://doi.org/10.1073/pnas.2304748120>.
- Jin, Youngjin, Eugene Jang, Jian Cui, Jin-Woo Chung, Yongjae Lee, and Seungwon Shin. 2023. "DarkBERT: A Language Model for the Dark Side of the Internet." <https://doi.org/10.48550/ARXIV.2305.08596>.
- Jing, Bowen, Bonnie Berger, and Tommi Jaakkola. 2024. "AlphaFold Meets Flow Matching for Generating Protein Ensembles." arXiv. <http://arxiv.org/abs/2402.04845>.
- Kaddour, Jean, Joshua Harris, Maximilian Mozes, Herbie Bradley, Roberta Raileanu, and Robert McHardy. 2023. "Challenges and Applications of Large Language Models." <https://doi.org/10.48550/ARXIV.2307.10169>.
- Kallini, Julie, Isabel Papadimitriou, Richard Futrell, Kyle Mahowald, and Christopher Potts. 2024. "Mission: Impossible Language Models." arXiv. <http://arxiv.org/abs/2401>.

References

- 06416.
- Kanjee, Zahir, Byron Crowe, and Adam Rodman. 2023. "Accuracy of a Generative Artificial Intelligence Model in a Complex Diagnostic Challenge." *JAMA* 330 (1): 78. <https://doi.org/10.1001/jama.2023.8288>.
- Kaufman, Jaycee M., Anirudh Thommandram, and Yan Fossat. 2023. "Acoustic Analysis and Prediction of Type 2 Diabetes Mellitus Using Smartphone-Recorded Voice Segments." *Mayo Clinic Proceedings: Digital Health* 1 (4): 534–44. <https://doi.org/10.1016/j.mcpdig.2023.08.005>.
- Killock, David. 2020. "AI Outperforms Radiologists in Mammographic Screening." *Nature Reviews Clinical Oncology* 17 (3): 134–34. <https://doi.org/10.1038/s41571-020-0329-7>.
- Kim, Yubin, Xuhai Xu, Daniel McDuff, Cynthia Breazeal, and Hae Won Park. 2024. "Health-LLM: Large Language Models for Health Prediction via Wearable Sensor Data." arXiv. <http://arxiv.org/abs/2401.06866>.
- Kung, Tiffany H., Morgan Cheatham, ChatGPT, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, et al. 2022. "Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models." Preprint. Medical Education. <https://doi.org/10.1101/2022.12.19.22283643>.
- Larson, Erik J. 2021a. *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*. Cambridge, Massachusetts London, England: The Belknap Press of Harvard University Press.
- . 2021b. *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.
- Lee, Cecilia S., Doug M. Baughman, and Aaron Y. Lee. 2017. "Deep Learning Is Effective for Classifying

- Normal Versus Age-Related Macular Degeneration OCT Images.” *Ophthalmology Retina* 1 (4): 322–27. <https://doi.org/10.1016/j.oret.2016.12.009>.
- Lee, Peter, Sebastien Bubeck, and Joseph Petro. 2023. “Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine.” Edited by Jeffrey M. Drazen, Isaac S. Kohane, and Tze-Yun Leong. *New England Journal of Medicine* 388 (13): 1233–39. <https://doi.org/10.1056/NEJMSr2214184>.
- Lee, Peter, Carey Goldberg, and Isaac Kohane. 2023. *The AI Revolution in Medicine: GPT-4 and Beyond*. 1st ed. Hoboken: Pearson.
- Leivada, Evelina, Elliot Murphy, and Gary Marcus. 2022. “DALL-E 2 Fails to Reliably Capture Common Syntactic Processes.” arXiv. <http://arxiv.org/abs/2210.12889>.
- Lenat, Doug, and Gary Marcus. 2023. “Getting from Generative AI to Trustworthy AI: What LLMs Might Learn from Cyc.” arXiv. <http://arxiv.org/abs/2308.04445>.
- Li, Luchang, Sheng Qian, Jie Lu, Lunxi Yuan, Rui Wang, and Qin Xie. 2024. “Transformer-Lite: High-Efficiency Deployment of Large Language Models on Mobile Phone GPUs.” arXiv. <http://arxiv.org/abs/2403.20041>.
- Liu, Nelson F., Tianyi Zhang, and Percy Liang. 2023. “Evaluating Verifiability in Generative Search Engines.” arXiv. <http://arxiv.org/abs/2304.09848>.
- Liu, Xiao, Hao Yu, Hanchen Zhang, Yifan Xu, Xuanyu Lei, Hanyu Lai, Yu Gu, et al. 2023. “AgentBench: Evaluating LLMs as Agents.” <https://doi.org/10.48550/ARXIV.2308.03688>.
- Liu, Zechun, Changsheng Zhao, Forrest Iandola, Chen Lai, Yuandong Tian, Igor Fedorov, Yunyang Xiong, et al. 2024. “MobileLLM: Optimizing Sub-Billion Parameter Language Models for On-Device Use Cases.” arXiv. <http://arxiv.org/abs/2402.14905>.

References

- Liu, Zhuoran, Leqi Zou, Xuan Zou, Caihua Wang, Biao Zhang, Da Tang, Bolin Zhu, et al. 2022. “Monolith: Real Time Recommendation System With Collisionless Embedding Table.” arXiv. <https://doi.org/10.48550/ARXIV.2209.07663>.
- Lu, Chris, Cong Lu, Robert Tjarko Lange, Jakob Foerster, Jeff Clune, and David Ha. 2024. “The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery.” arXiv. <http://arxiv.org/abs/2408.06292>.
- Luo, Renqian, Liai Sun, Yingce Xia, Tao Qin, Sheng Zhang, Hoifung Poon, and Tie-Yan Liu. 2022. “BioGPT: Generative Pre-Trained Transformer for Biomedical Text Generation and Mining.” *Briefings in Bioinformatics* 23 (6): bbac409. <https://doi.org/10.1093/bib/bbac409>.
- Lutsker, Guy, Gal Sapir, Anastasia Godneva, Smadar Shilo, Jerry R. Greenfield, Dorit Samocha-Bonet, Shie Manor, et al. 2024. “From Glucose Patterns to Health Outcomes: A Generalizable Foundation Model for Continuous Glucose Monitor Data Analysis.” arXiv. <http://arxiv.org/abs/2408.11876>.
- Ma, Xiao, Swaroop Mishra, Ahmad Beirami, Alex Beutel, and Jilin Chen. 2023. “Let’s Do a Thought Experiment: Using Counterfactuals to Improve Moral Reasoning.” <https://doi.org/10.48550/ARXIV.2306.14308>.
- Mahowald, Kyle, Anna A. Ivanova, Idan A. Blank, Nancy Kanwisher, Joshua B. Tenenbaum, and Evelina Fedorenko. 2023. “Dissociating Language and Thought in Large Language Models: A Cognitive Perspective.” arXiv. <http://arxiv.org/abs/2301.06627>.
- Manathunga, Supun, and Isuru Hettigoda. 2023. “Aligning Large Language Models for Clinical Tasks.” arXiv. <http://arxiv.org/abs/2309.02884>.
- McDuff, Daniel, Mike Schaekermann, Tao Tu, Anil Palepu, Amy Wang, Jake Garrison, Karan Singhal, et al. 2023. “Towards

- Accurate Differential Diagnosis with Large Language Models.” arXiv. <http://arxiv.org/abs/2312.00164>.
- Meskó, Bertalan. 2023. “Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial.” *Journal of Medical Internet Research* 25 (October): e50638. <https://doi.org/10.2196/50638>.
- Meskó, Bertalan, and Eric J. Topol. 2023. “The Imperative for Regulatory Oversight of Large Language Models (or Generative AI) in Healthcare.” *Npj Digital Medicine* 6 (1): 120. <https://doi.org/10.1038/s41746-023-00873-0>.
- Millière, Raphaël, and Cameron Buckner. 2024. “A Philosophical Introduction to Language Models – Part I: Continuity With Classic Debates.” arXiv. <http://arxiv.org/abs/2401.03910>.
- Narayanan, Arvind, and Sayash Kapoor. 2024. *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference*. Princeton Oxford: Princeton University Press.
- Nori, Harsha, Nicholas King, Scott Mayer McKinney, Dean Carignan, and Eric Horvitz. 2023. “Capabilities of GPT-4 on Medical Challenge Problems.” <https://doi.org/10.48550/ARXIV.2303.13375>.
- Oh, Hamilton Se-Hwee, Jarod Rutledge, Daniel Nachun, Róbert Pálovics, Olamide Abiose, Patricia Moran-Losada, Divya Channappa, et al. 2023. “Organ Aging Signatures in the Plasma Proteome Track Health and Disease.” *Nature* 624 (7990): 164–72. <https://doi.org/10.1038/s41586-023-06802-1>.
- Oren, Yonatan, Nicole Meister, Niladri Chatterji, Faisal Ladhak, and Tatsunori B. Hashimoto. 2023. “Proving Test Set Contamination in Black Box Language Models.” arXiv. <http://arxiv.org/abs/2310.17623>.
- Pei, Gan, Jiangning Zhang, Menghan Hu, Guangtao Zhai,

References

- Chengjie Wang, Zhenyu Zhang, Jian Yang, Chunhua Shen, and Dacheng Tao. 2024. “Deepfake Generation and Detection: A Benchmark and Survey.” arXiv. <http://arxiv.org/abs/2403.17881>.
- Qian, Cheng, Xinran Zhao, and Sherry Tongshuang Wu. 2023. “”Merge Conflicts!” Exploring the Impacts of External Distractors to Parametric Knowledge Graphs.” arXiv. <http://arxiv.org/abs/2309.08594>.
- Qiu, Pengcheng, Chaoyi Wu, Xiaoman Zhang, Weixiong Lin, Haicheng Wang, Ya Zhang, Yanfeng Wang, and Weidi Xie. 2024. “Towards Building Multilingual Language Model for Medicine.” arXiv. <http://arxiv.org/abs/2402.13963>.
- Raji, Inioluwa Deborah, I. Elizabeth Kumar, Aaron Horowitz, and Andrew D. Selbst. 2022. “The Fallacy of AI Functionality.” In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 959–72. <https://doi.org/10.1145/3531146.3533158>.
- Rao, Arya, Michael Pang, John Kim, Meghana Kamineni, Winston Lie, Anoop K Prasad, Adam Landman, Keith Dreyer, and Marc D Succi. 2023. “Assessing the Utility of ChatGPT Throughout the Entire Clinical Workflow: Development and Usability Study.” *Journal of Medical Internet Research* 25 (August): e48659. <https://doi.org/10.2196/48659>.
- Romera-Paredes, Bernardino, Mohammadamin Barekatain, Alexander Novikov, Matej Balog, M. Pawan Kumar, Emilien Dupont, Francisco J. R. Ruiz, et al. 2024. “Mathematical Discoveries from Program Search with Large Language Models.” *Nature* 625 (7995): 468–75. <https://doi.org/10.1038/s41586-023-06924-6>.
- Röttger, Paul, Valentin Hofmann, Valentina Pyatkin, Musashi Hinck, Hannah Rose Kirk, Hinrich Schütze, and Dirk Hovy. 2024. “Political Compass or Spinning Arrow? Towards More Meaningful Evaluations for Val-

- ues and Opinions in Large Language Models.” arXiv. <http://arxiv.org/abs/2402.16786>.
- Rozado, David. 2023. “The Political Biases of ChatGPT.” *Social Sciences* 12 (3): 148. <https://doi.org/10.3390/socsci12030148>.
- . 2024. “The Political Preferences of LLMs.” arXiv. <http://arxiv.org/abs/2402.01789>.
- Saab, Khaled, Tao Tu, Wei-Hung Weng, Ryutaro Tanno, David Stutz, Ellery Wulczyn, Fan Zhang, et al. 2024. “Capabilities of Gemini Models in Medicine.” arXiv. <http://arxiv.org/abs/2404.18416>.
- Sastry, Girish, Lennart Heim, Haydn Belfield, Markus Anderljung, Miles Brundage, Julian Hazell, Cullen O’Keefe, et al. 2024. “Computing Power and the Governance of Artificial Intelligence.” arXiv. <http://arxiv.org/abs/2402.08797>.
- Shumailov, Ilia, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicolas Papernot, and Ross Anderson. 2023. “The Curse of Recursion: Training on Generated Data Makes Models Forget.” arXiv. <http://arxiv.org/abs/2305.17493>.
- Singhal, Karan, Shekoofeh Azizi, Tao Tu, S. Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, et al. 2023. “Large Language Models Encode Clinical Knowledge.” *Nature* 620 (7972): 172–80. <https://doi.org/10.1038/s41586-023-06291-2>.
- Sun, Jie, Qun-Xi Dong, San-Wang Wang, Yong-Bo Zheng, Xiao-Xing Liu, Tang-Sheng Lu, Kai Yuan, et al. 2023. “Artificial Intelligence in Psychiatry Research, Diagnosis, and Therapy.” *Asian Journal of Psychiatry* 87 (September): 103705. <https://doi.org/10.1016/j.ajp.2023.103705>.
- Tian, Yuzhang, Jianbo Zhao, Haoyu Dong, Junyu Xiong, Shiyu Xia, Mengyu Zhou, Yun Lin, et al. 2024. “SpreadsheetLLM: Encoding Spreadsheets for Large Language Models.” arXiv. <http://arxiv.org/abs/2407.09025>.

References

- Tu, Tao, Anil Palepu, Mike Schaekermann, Khaled Saab, Jan Freyberg, Ryutaro Tanno, Amy Wang, et al. 2024. "Towards Conversational Diagnostic AI." <https://doi.org/10.48550/ARXIV.2401.05654>.
- Udandaraao, Vishaal, Ameya Prabhu, Adhiraj Ghosh, Yash Sharma, Philip H. S. Torr, Adel Bibi, Samuel Albanie, and Matthias Bethge. 2024. "No "Zero-Shot" Without Exponential Data: Pretraining Concept Frequency Determines Multimodal Model Performance." arXiv. <http://arxiv.org/abs/2404.04125>.
- Villalobos, Pablo, Jaime Sevilla, Lennart Heim, Tamay Besiroglu, Marius Hobbhahn, and Anson Ho. 2022. "Will We Run Out of Data? An Analysis of the Limits of Scaling Datasets in Machine Learning." arXiv. <http://arxiv.org/abs/2211.04325>.
- Wang, Chengliang, Xinrun Chen, Haojian Ning, and Shiyong Li. 2023. "SAM-OCTA: A Fine-Tuning Strategy for Applying Foundation Model to OCTA Image Segmentation Tasks." arXiv. <http://arxiv.org/abs/2309.11758>.
- Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models." arXiv. <http://arxiv.org/abs/2201.11903>.
- Wei, Jerry, Chengrun Yang, Xinying Song, Yifeng Lu, Nathan Hu, Dustin Tran, Daiyi Peng, et al. 2024. "Long-Form Factuality in Large Language Models." arXiv. <http://arxiv.org/abs/2403.18802>.
- Weiss, Roy, Daniel Ayzenshteyn, Guy Amit, and Yisroel Mirsky. 2024. "What Was Your Prompt? A Remote Keylogging Attack on AI Assistants." arXiv. <http://arxiv.org/abs/2403.09751>.
- Wendler, Chris, Veniamin Veselovsky, Giovanni Monea, and

- Robert West. 2024. “Do Llamas Work in English? On the Latent Language of Multilingual Transformers.” arXiv. <https://arxiv.org/abs/2402.10588>.
- Wornow, Michael, Yizhe Xu, Rahul Thapa, Birju Patel, Ethan Steinberg, Scott Fleming, Michael A. Pfeffer, Jason Fries, and Nigam H. Shah. 2023. “The Shaky Foundations of Large Language Models and Foundation Models for Electronic Health Records.” *Npj Digital Medicine* 6 (1): 135. <https://doi.org/10.1038/s41746-023-00879-8>.
- Yiu, Eunice, Eliza Kosoy, and Alison Gopnik. 2023. “Transmission Versus Truth, Imitation Versus Innovation: What Children Can Do That Large Language and Language-and-Vision Models Cannot (Yet).” *Perspectives on Psychological Science*, October, 17456916231201401. <https://doi.org/10.1177/17456916231201401>.
- Yu, Feiyang, Mark Endo, Rayan Krishnan, Ian Pan, Andy Tsai, Eduardo Pontes Reis, Eduardo Kaiser Ururahy Nunes Fonseca, et al. 2023. “Evaluating Progress in Automatic Chest X-Ray Radiology Report Generation.” *Patterns* 4 (9): 100802. <https://doi.org/10.1016/j.patter.2023.100802>.
- Zaleski, Amanda L, Rachel Berkowsky, Kelly Jean Thomas Craig, and Linda S Pescatello. 2024. “Comprehensiveness, Accuracy, and Readability of Exercise Recommendations Provided by an AI-Based Chatbot: Mixed Methods Study.” *JMIR Medical Education* 10 (January): e51308. <https://doi.org/10.2196/51308>.
- Zhao, Theodore, Yu Gu, Jianwei Yang, Naoto Usuyama, Ho Hin Lee, Sid Kiblawi, Tristan Naumann, et al. 2024. “A Foundation Model for Joint Segmentation, Detection and Recognition of Biomedical Objects Across Nine Modalities.” *Nature Methods*, November. <https://doi.org/10.1038/s41592-024-02499-w>.
- Zhao, Zihao, Yuxiao Liu, Han Wu, Yonghao Li, Sheng Wang, Lin

References

- Teng, Disheng Liu, et al. 2023. “CLIP in Medical Imaging: A Comprehensive Survey.” arXiv. <http://arxiv.org/abs/2312.07353>.
- Zheng, Huaixiu Steven, Swaroop Mishra, Hugh Zhang, Xinyun Chen, Minmin Chen, Azade Nova, Le Hou, et al. 2024. “NATURAL PLAN: Benchmarking LLMs on Natural Language Planning.” arXiv. <http://arxiv.org/abs/2406.04520>.
- Zhou, Shuyan, Frank F. Xu, Hao Zhu, Xuhui Zhou, Robert Lo, Abishek Sridhar, Xianyi Cheng, et al. 2023. “WebArena: A Realistic Web Environment for Building Autonomous Agents.” arXiv. <http://arxiv.org/abs/2307.13854>.