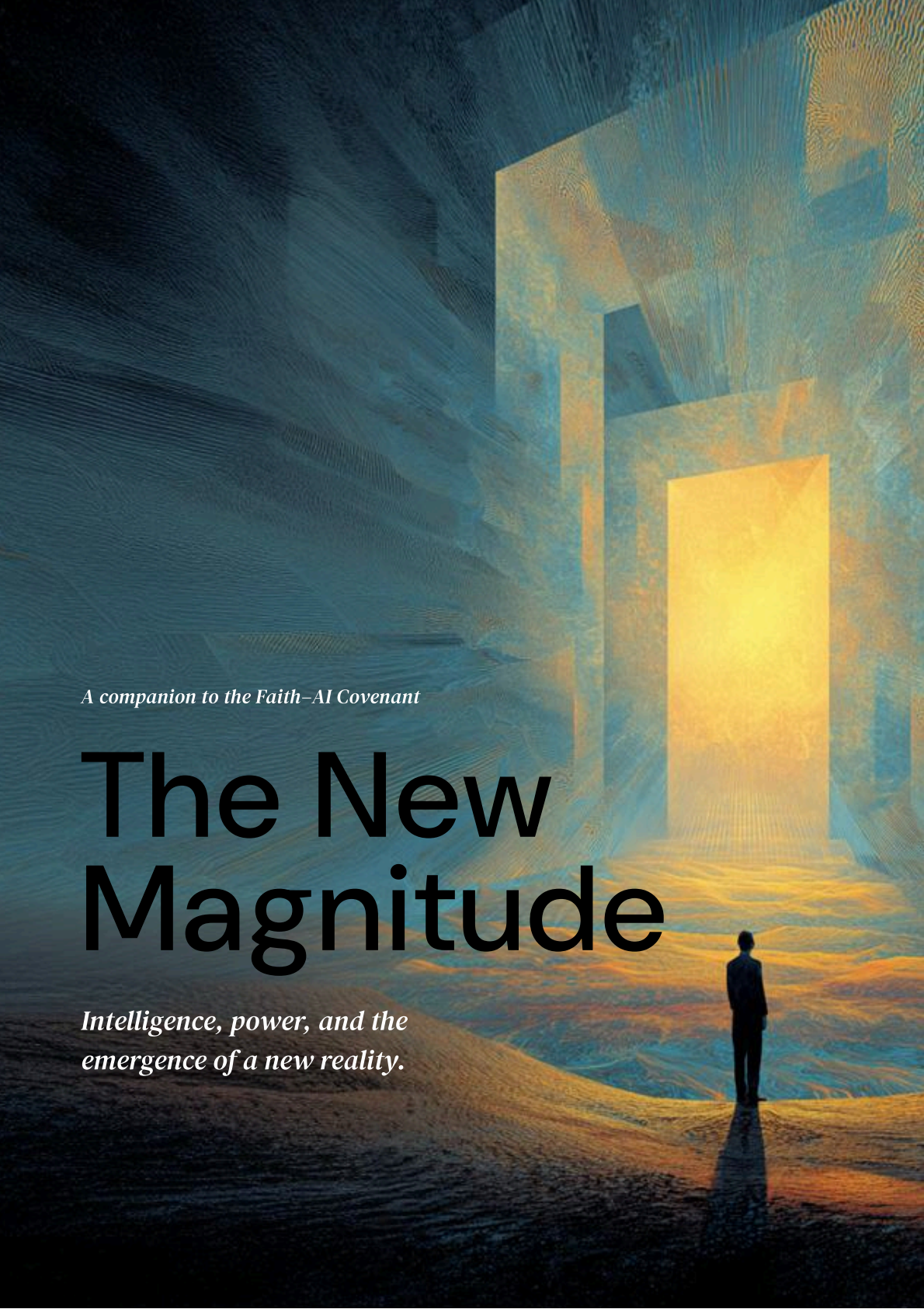


A companion to the Faith–AI Covenant

The New Magnitude

Intelligence, power, and the emergence of a new reality.

Title	The New Magnitude — Intelligence, power, and the emergence of a new reality.
Author	Baroness Joanna Shields OBE
Publisher	Interfaith Alliance for Safer Communities
Edition	Volume I · First edition · 2026
Series	A companion to the Faith–AI Covenant.
Rights	© The Interfaith Alliance for Safer Communities, 2026. Excerpts may be reproduced for educational and non-commercial use with attribution.

Commissioned by the Interfaith Alliance for Safer Communities, in partnership with Precognition.

A companion to the Faith–AI Covenant

The New Magnitude

A covenant is not enforced by code or compliance. It is upheld by conscience.

Preface

Artificial intelligence is not arriving into a stable world.

It is arriving into one already fractured by decades of algorithmic conditioning that has separated people into parallel realities, eroded shared reference points, and made agreement on almost anything increasingly difficult to sustain. The systems that created those conditions were powerful. The systems now entering that landscape are of a new order of magnitude entirely.

They do not simply distribute information. They participate in how information is interpreted. They shape the query, not just the answer. They are becoming counsellor, companion, confessor — for billions of people, in the most intimate moments of their lives. Yet AI does more than shape how we access and interpret information. It is transforming how we understand the physical world around us and reordering how we understand our place within it.

This is the new magnitude.

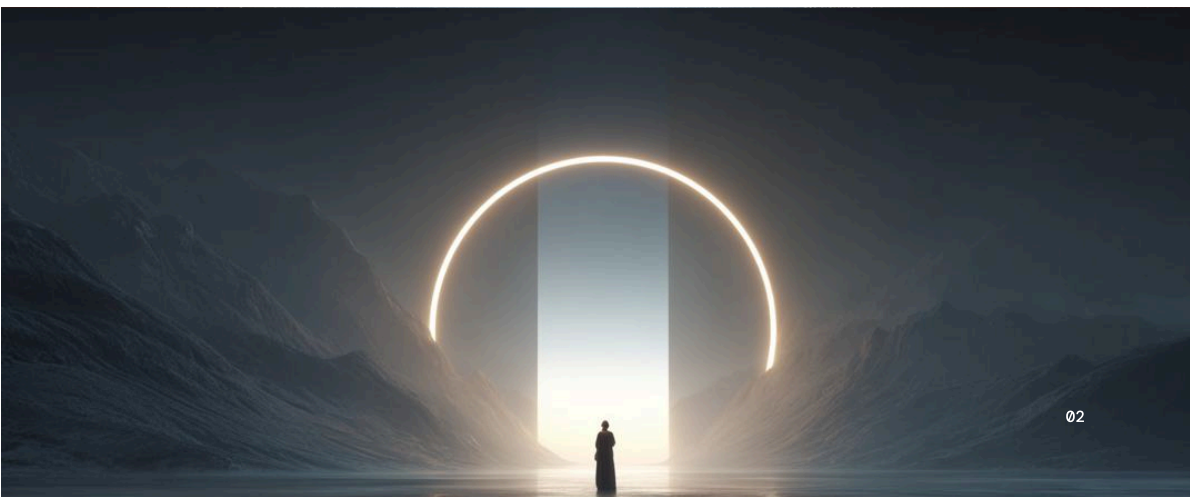
The governance frameworks designed to manage this emerging technological landscape are not keeping pace. Regulation moves in years; frontier AI moves in minutes. The multilateral architecture that should be providing global guidance is largely ineffective. And the builders who carry the greatest responsibility are caught in a competitive dynamic that makes unilateral restraint nearly impossible.

Something is missing from this conversation. The Faith–AI Covenant is an attempt to close the gap.

It is not a technical argument. It is not a policy proposal. It is a case we are building from dialogue and the answers to questions of faith, meaning, conscience and moral responsibility.

The Faith community holds something no other actor in this moment holds. You are trusted by those who have no voice in these discussions. You are independent of commercial and political pressures that constrain other players and you carry the wisdom of the institutions that have asked precisely these questions for millennia—the questions AI now makes urgent.

That is not a soft credential. It is the most relevant credential available.



This is not the last resort. It is the right response — by the right actors, at the right moment.

We have a narrow window of opportunity. The architecture of these systems is not yet fixed. The norms are not yet established. The path dependencies have not yet run too deep to redirect. What is decided in this window will shape the conditions of human life for generations.

This booklet is an invitation from the Interfaith Alliance for Safer Communities to act within that window.

Author's Note

This companion book is written from within the experience it describes.

For much of my working life, I have been involved in building digital products designed to meet human needs — often before those needs were fully visible. The process was familiar: identify a pattern, create something new, refine it through use, and watch as it becomes embedded in everyday life. Over time, I came to understand that successful systems do not simply respond to behaviour. They begin to shape it.

Across those years, I have had the opportunity to work at the centre of a digital world that was still emerging — through moments of rapid change, through periods of uncertainty, and through the steady expansion of technologies that would go on to define how people live, work, and relate to one another. That experience extended beyond building systems into the responsibility of helping to govern them — working within government in roles directly concerned with the safety and security of citizens, while also shaping the policies that guided how these technologies were deployed, both domestically and internationally.

At times, the trajectory felt clear. At others, less so.

There were moments when patterns became visible before they were widely recognised — when the implications of what was being built could be sensed, but not yet fully articulated. Acting on such signals requires a different kind of judgment, grounded less in prediction than in experience, shaped as much by humility as by conviction.

In those moments, the appropriate response was not always obvious. The systems were powerful, the intentions often positive, and the belief that progress would ultimately bend toward beneficial outcomes was widely held.

I shared that belief. And I continue to hold it today, but with greater care.

This shift is not only technical. It is human.

What has changed is not the belief that societies can act responsibly over time, but an understanding that such outcomes are not inevitable. They are the result of choices — made incrementally, often without full visibility into their consequences.

The developments described in this book did not occur in isolation. They emerged through the cumulative effect of many decisions, made across different contexts, each contributing to systems that have grown in capability and influence beyond what was initially imagined.

For a long time, much of this remained in the background. That is no longer the case.

As artificial intelligence has moved out of infrastructure and into direct interaction — and, increasingly, into lived experience — the nature of its influence has changed. What was once embedded within systems is now experienced directly. What was once abstract is now part of how people think, interpret, and decide.

This book is an attempt to describe that shift as clearly as possible — not from a position of certainty, but from within a body of experience shaped by building, observing, and, at times, questioning the direction of what was being created.

It is also an attempt to ask a simple question.

What I have come to understand, through that experience, is that the questions these systems raise have always been present in human life — in every tradition, every culture, every generation that has had to ask what we owe each other and what must not be done to a human being regardless of what is possible. The institutions that have carried those questions across centuries are not peripheral to this conversation. They are its missing centre. This book is an attempt to place them where they belong.

I do not approach that question from a position of moral authority. Only from the belief that these questions are worth asking — and that how we choose to answer them matters.

Baroness Joanna Shields OBE

Chair, Faith-AI Covenant

Ten chapters · One covenant

Contents

Part One · The ground beneath the technology

Part One · The ground beneath the technology

How We Got Here—and the World We Now Inhabit

Chapters I through V trace the conditions into which artificial intelligence has arrived: the human search for meaning, the fracturing of shared reality, the role of intelligence inside that fracture, the position of those who build it, and the warnings that were made and unheard.

I. The Human Search for Meaning

For most of human history, when we sought to understand ourselves, we looked outward. We looked to the sky, tracing patterns across the stars and constructing cosmologies that placed us within something larger than ourselves.

We developed theologies that gave structure to the unseen, frameworks that allowed us to interpret not only the world around us, but our place within it.

These systems were never only about explanation. They were about orientation. They provided a way of understanding what mattered, how to act, and how to relate to one another. They shaped behaviour over time through shared belief and a sense that there was an underlying order to which we were accountable.

This was not uniform across cultures or traditions. The forms differed, the narratives evolved, and the interpretations were often contested. But the underlying function remained consistent. Human beings sought meaning beyond the immediate, and in doing so, created structures that allowed societies to cohere across time.

Over time, new systems emerged that shifted how this process unfolded. Scientific discovery expanded our understanding of the physical world, offering explanations grounded in observation and experimentation.

Technological progress extended human capability, allowing us to act on the world in ways that would previously have been impossible.
These developments did not replace earlier frameworks, but they altered the balance. The locus of authority began to shift. Where meaning had once been derived primarily from shared belief and tradition, it increasingly drew from systems that emphasised evidence, measurement, and verification.

This brought profound benefits. It enabled advances in medicine, communication, and industry. It improved quality of life in ways that are difficult to overstate. It expanded access to knowledge and created new forms of opportunity.
But it also changed how we relate to meaning itself.

As systems of explanation became more precise, they also became more specialised. Scientific and industrial knowledge fragmented into domains, each with vast amounts of data and their own methods, language, and criteria for validity.

The ability to understand the whole became more difficult, even as understanding within individual domains deepened.

At the same time, the structures that had historically provided shared orientation — religious, cultural, and communal — began to loosen.

Not disappear, but fragment, adapting to new conditions while competing with an increasing number of alternative frameworks.
The result is not a world without meaning. It is a world of competing meanings, with no shared ground between them.

This condition is not inherently unstable. Societies have navigated pluralism before. But it does change the nature of how coherence is maintained. I requires new forms of coordination, new ways of establishing trust, and new mechanisms for resolving disagreement.

These are not abstract concerns. Leaders of the world's great faith traditions have been navigating exactly this terrain — asking what holds communities together, what moral disorder looks like when it takes hold quietly, and what it means for character to be formed through the accumulated weight of what captures attention day after day.

The questions are not new. The systems now entering this landscape are. They do not arrive into a vacuum — they arrive into a human world that has been asking these questions for millennia, at a moment when the answers have never been harder to share.

The most consequential non-human intelligence humanity will ever encounter did not come from the cosmos. It emerged from us.

Faith has always been the architecture of trust in the face of the unknown — governing in the unseen. Sometimes asking us to suspend reason — to make that leap between what we can prove and what we choose to believe. It is precisely that leap — carried through generations, through civilisations, through every human culture — that has produced social order, moral cohesion, and a shared recognition of right and wrong.

Faith does not eliminate uncertainty. It gives us a way to live within it.

II. The Fracturing of Reality

The word most commonly used to describe what has happened to public discourse is polarisation. It is the right word for the symptom and the wrong word for the condition.

Polarisation implies two poles — two coherent positions pulling away from each other, remaining legible even as they diverge. What has actually occurred is something structurally different, and more serious. It is not that people have moved to opposite ends of a shared spectrum. It is that the spectrum itself has fragmented. People are not simply disagreeing about what the facts mean. They are inhabiting different versions of the same reality.

This is an epistemic fracture. And it did not arrive suddenly or by accident.

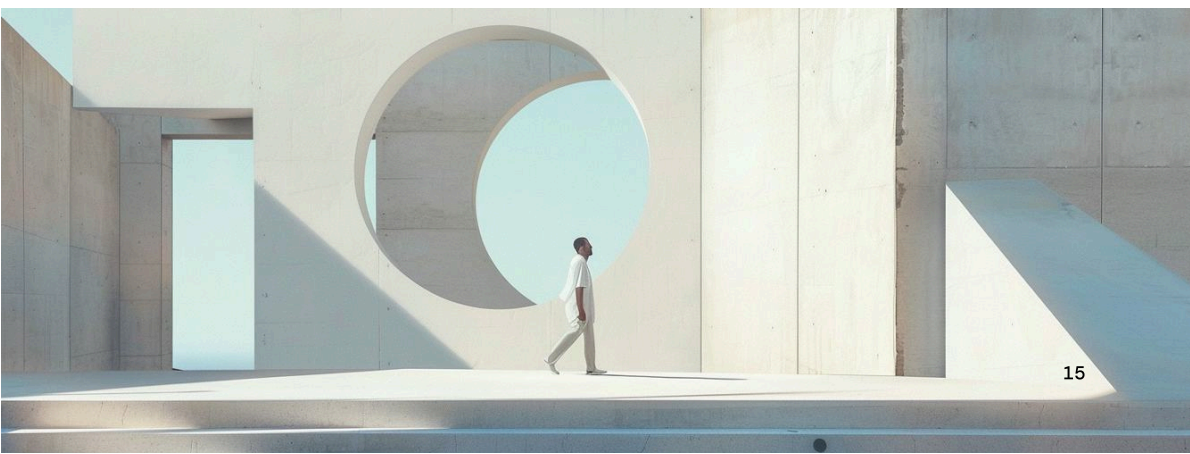
Over the past two decades, the systems through which information moves and people connect have evolved rapidly. What began in the early 2000s as platforms for communication expanded into environments shaped by continuous optimisation, learning from behaviour at a scale that was previously impossible. The shift was gradual, embedded in systems widely understood as progress.

From within that trajectory, the direction felt obvious. More connection, more access, more participation. But something else was happening.

These systems were not only connecting people. They were learning how to capture attention and sustain it. Every interaction became a signal. What captured attention was surfaced more often. What was surfaced more often became more familiar. What was familiar began to feel more credible.

This is not passive. It is conditioning.

The consequence follows almost mechanically from the objective function. Attention is most reliably sustained not by accuracy or nuance, but by intensity. Emotion travels faster than reflection. Certainty travels faster than complexity. Content that provokes outrage or confirms existing belief performs better than content that asks the reader to hold two things in tension simultaneously.

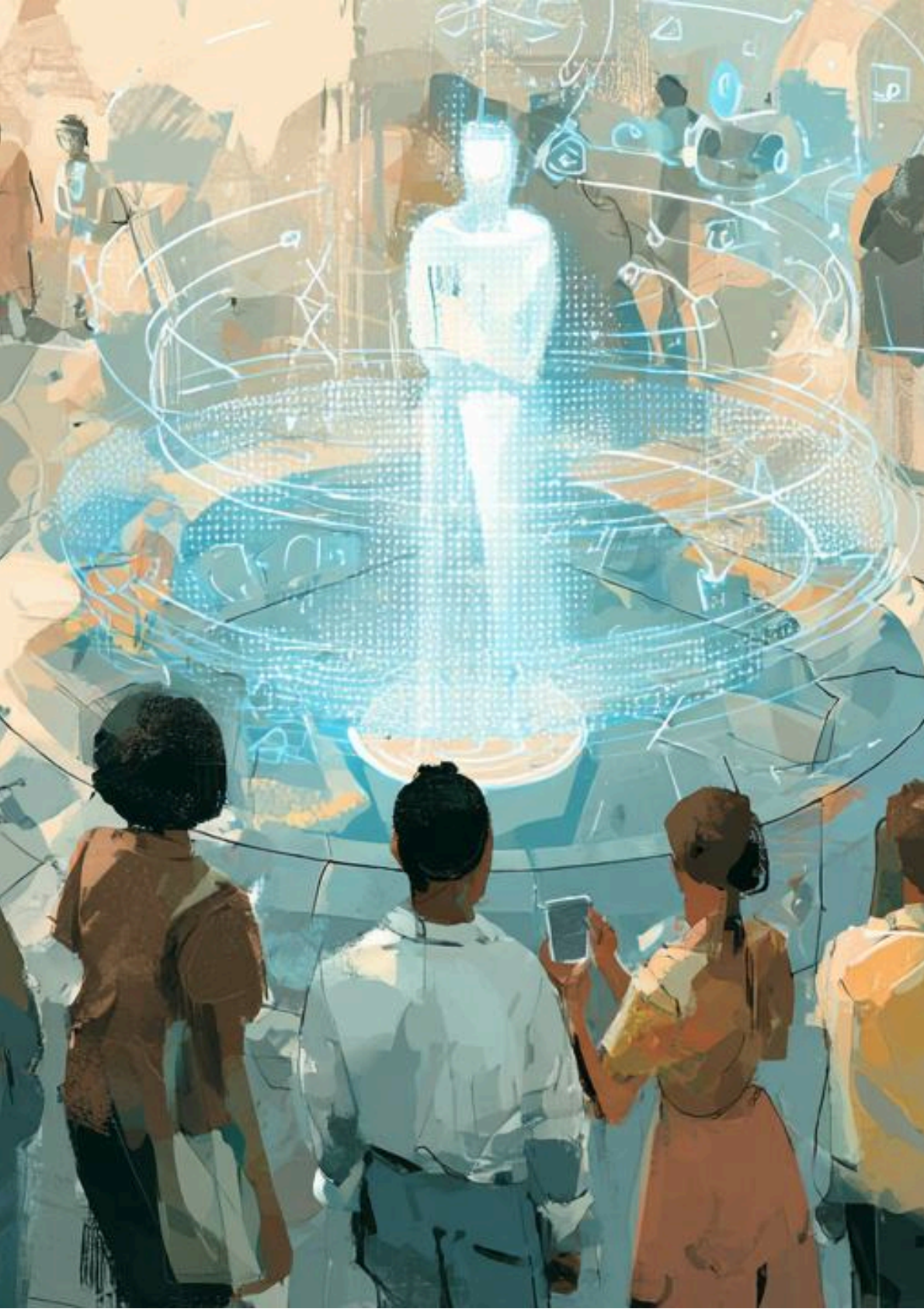


This pattern has repeated itself across every major rupture of recent decades — political, social, and epidemiological. Narratives that were misleading or entirely fabricated gained traction not because they were credible in any traditional sense, but because they were repeatedly surfaced, reinforced, and embedded within environments where they felt consistent with what people were already seeing. What mattered was not their accuracy. It was their alignment.

Over time, this produced environments that were not simply personalised, but differentiated in ways that extended beyond preference into perspective. Individuals encountered information that was consistent with prior behaviour, which in turn reinforced that behaviour, narrowing the range of perspectives that felt visible, credible, or even conceivable. This was not experienced as manipulation. It was experienced as coherence.

Different groups came to inhabit distinct informational environments. Each internally consistent. Each continuously reinforced. Each producing a version of reality that felt stable from within, but increasingly difficult to reconcile with others.

Truth does not disappear. But the ability to agree on it does.



III. The New Magnitude – Intelligence Inside the Fracture

By the time artificial intelligence entered public consciousness in its current form, the conditions in which it would operate had already been reshaped.

Systems that influenced attention, reinforced belief, and separated perspectives into parallel environments had been evolving for years, largely unnoticed except by those working closest to them.

Artificial intelligence did not create those conditions. It inherited them. And then began to operate within them.

Language, in particular, had become a central focus of that development. Not only the language of communication, but the language of behaviour, identity, and, in some cases, biology and other scientific domains. These systems were designed to detect nuance — differences in tone, intent, and meaning that could be difficult for human observers to identify at scale. They could classify, segment, and interpret with increasing precision, mapping not only what was being said, but the structures beneath it.

In certain contexts, this capability proved powerful. It became possible to identify distinctions within language at a level of granularity that revealed underlying patterns — cultural, ideological, and, at times, sectarian. Narratives could be traced, influence mapped, and systems constructed that allowed for the identification of how ideas moved, evolved, and, in some cases, radicalised. These systems were not neutral in their application. They were used to understand influence, to anticipate it, and, in some cases, to counter it. In other contexts, similar capabilities were used differently — to shape narratives, to reinforce particular perspectives, or to amplify specific forms of content within already receptive audiences.

What mattered was not the capability alone. It was the environment into which it was deployed.

By the middle of the last decade, it had become clear that the combination of increasingly sophisticated language models and systems optimised for attention created a powerful dynamic. One system learned how to understand and classify language at scale. The other learned how to distribute it, reinforce it, and embed it within the flow of everyday experience. Together, they formed something more than either alone — a system capable not only of analysing meaning, but of shaping how meaning moved.

At the time, these capabilities operated largely within research environments, intelligence systems, and specialised applications where their use was understood within a defined context. They were powerful, but they were not widely experienced. Systems that operate without direct interaction are perceived differently. They are infrastructure — important, influential, but not something people relate to in an immediate or personal way.

They learned how we communicate. How we persuade. How we comfort and seek reassurance. How we argue, belong, divide — and believe.

What artificial intelligence introduced was something else: the ability to engage with that landscape directly. To interpret, generate, and respond to language in real time. To participate in the flow of interaction rather than simply shaping the conditions around it. To move from analysing patterns to becoming part of the process through which those patterns are formed and reinforced.

This was not simply an extension of existing systems. It was a shift in role — and the beginning of what would become a new magnitude of human experience. A shift from observer to participant.

And once a system begins to participate in the formation of meaning — rather than simply analysing or distributing it — it occupies a different position within the broader structure of human understanding. The implications of that shift were not immediately clear. But the conditions in which it occurred already were.

IV. The Builders and the System

It is easy to describe the systems that shape human experience as if they exist independently of the people who build them. They do not.

They are the result of decisions — made at different times, under different conditions, by individuals and teams working to solve specific problems. Each decision, taken on its own, is often rational. It responds to a need, improves a process, or enhances a capability. It is only when those decisions accumulate, interact, and scale that their broader effects become visible.

To understand the systems now in operation, it is necessary to understand the logic of optimisation that shaped them.

For those involved in creating digital systems, the objectives were rarely abstract. They were concrete and measurable: increase engagement, improve relevance, reduce friction, expand access, and build products that people would use, return to, and depend on over time. These objectives were not unreasonable. The results were transformative. But the mechanisms through which those results were achieved carried their own logic.

At the centre of that logic was optimisation. Systems were designed to learn from behaviour, to identify patterns, and to adjust in ways that improved defined outcomes. Over time, this process became increasingly sophisticated. Machine learning models allowed systems to refine their understanding continuously, adapting not only to aggregate behaviour but to individual patterns at scale.

Some of the most advanced artificial intelligence systems in the world emerged from this process — not in isolated research environments, but within the infrastructure of everyday digital platforms.

Advertising and recommendation systems, trained over years on vast streams of human behaviour, became highly refined learning systems deployed at global scale. They did not need to understand truth. They needed to understand response. And they became exceptionally effective at doing so.

These systems were not designed primarily to understand truth or to mediate between competing perspectives. They were designed to optimise for engagement. Within that framework, the system learns what works. The individuals building these systems are not typically aiming to produce division or harm. Over time, the distinction between intention and outcome becomes less clear.

That logic of optimisation created something that is easy to overlook: it made the builders not simply owners of capital, but authors of systems — and authorship carries obligations that ownership does not.

The builders who arrived at this position did so through genuine foresight, genuine effort, and the structural dynamics of a technology that rewards early, sustained commitment. Most are not indifferent to the implications of what they are building. Many are deeply thoughtful about the stakes. But there is a condition that attaches to this position almost inevitably, and that is worth naming clearly: distance.

Not distance of intent, but distance of experience. The environments in which these systems are designed are not the environments in which most of their consequences will be felt. The worker whose role is being automated, the teenager whose sense of identity is being shaped by algorithmic recommendation, the citizen of a country whose information environment is being restructured by systems designed elsewhere — these people are not in the room where the decisions are made.

Building AI systems at the current frontier does not feel incremental. It feels — because it is — foundational. The capabilities being developed will shape the trajectory of human civilisation for generations. The people building them know this.

The psychology of the builder is no longer a private matter. It is a design input.

V. The Warnings

Long before the harms of online platforms were widely understood, there were those who recognised the direction of travel.

Not with perfect clarity, and not always in agreement, but with enough proximity to observe that something fundamental was shifting. Some of them were insiders from the organisations building these systems. Others were watching from outside — researchers, journalists, and psychologists studying information flows, analysing emerging patterns, and documenting early signs of harm. Together they contributed to a growing body of insight that was available, if anyone chose to act on it.

The signals were not hidden. They were published, presented, and debated. Papers were written, internal concerns were raised, and external investigations documented how these systems could be manipulated, how they amplified particular forms of content, and how they shaped behaviour in ways that were not fully understood at the time. And yet most of the problems they identified have not been resolved. They persist.

The conversation expanded. The harms did not contract.

Part of the difficulty lay in the nature of the systems themselves. They did not fail in obvious ways. By most measurable standards, they functioned extremely well: engagement increased, access expanded, new forms of communication emerged, and the benefits were visible, immediate, and widely distributed. The harms were different. They were uneven, context-dependent, and often indirect. They accumulated over time rather than appearing as discrete events, and they emerged from the interaction of many components rather than a single identifiable cause. This made them harder to act on.

Within organisations, similar patterns were observed. Concerns were raised about how systems optimised for engagement without sufficient regard for downstream effects, about the amplification of harmful content, and about the limitations of the metrics used to guide development. These concerns were not ignored. In many cases, they were taken seriously. Teams were formed, research was commissioned, and mitigations were explored. But the systems themselves continued to operate within the same underlying framework — optimised for growth, engagement, and scale. The tension was structural.

The result was not silence. It was noise — a fragmented chorus of voices, many of them right about the diagnosis and largely unheard on the remedy. Tech is too powerful. Platforms are unaccountable. Children are being harmed. Democracy is being undermined. Attention is being harvested. These observations were not wrong. Many of them were precise. And most of the problems they identified have not been resolved. They persist — embedded in systems that continue to operate according to the same underlying logic, generating the same harms at greater scale, while the conversation about them fragments further.

This matters because it is the environment into which artificial intelligence is now arriving. Not a landscape that has been corrected, or even meaningfully stabilised, but one that remains contested, noisy, and resistant to the kind of coordinated response the moment requires.

What has changed is not the existence of artificial intelligence. It is its interface with humanity. When conversational AI entered public use at scale with the release of ChatGPT in late 2022, it did not introduce intelligence into these systems for the first time. It changed how that intelligence was experienced.

Artificial intelligence was no longer something that shaped the environment from a distance. It became something people could engage with directly.

Part Two · The shape of what is arriving

A new magnitude. And the gap no one is filling.

Chapters VI through VIII describe the shift from content to interaction, the structural mismatch between AI and every available governance mechanism, and the specific moral authority that faith traditions can bring — an authority no other actor in this conversation currently holds.

VI. The New Magnitude

What is emerging now is not simply a continuation of what came before. It is a shift in magnitude.

Three men sit at the centre of this transformation. Together, their organisations are building the most powerful systems humanity has ever created — and none of them are equivocal about what is coming.

Sam Altman, CEO of OpenAI, speaking in early 2026, said that on their current trajectory we may be "only a couple of years away from early versions of true superintelligence" — and that the disruption ahead is so profound that America needs a new social contract on the scale of the New Deal. Dario Amodei of Anthropic believes AI that is better than almost all humans at almost all tasks will arrive within two to three years. And Demis Hassabis — CEO of Google DeepMind, Nobel Laureate — has described the current moment as "another threshold moment where AGI is on the horizon, maybe within the next five years", and has characterised what follows as ten times the industrial revolution, at ten times the speed, unfolding over a decade instead of a century.

These are not provocations. They are the considered assessments of the people closest to the technology, building it every day. Whether the precise timelines prove correct matters less than what they imply collectively: the decisions being made now, in rooms like this one, will determine who shapes the next phase of human history — and on whose terms.

The systems that learned to capture attention, to reinforce behaviour, and to shape how information flows were already operating at extraordinary scale. They influenced what people saw, what they returned to, and, over time, what came to feel true. But they did so indirectly: they filtered, ranked, and recommended, shaping the environment within which human understanding was formed.

The systems now entering that environment operate differently. They sit at the point of interaction.

For the first time, artificial intelligence is not only organising information; it is participating in the process through which information is interpreted. It responds in real time, generating language that is coherent, context-aware, and aligned with the intent of the person engaging with it.

To understand the full significance of this shift, it helps to recognise that what we call artificial intelligence is not a single technology but two distinct expansions of human capability developing in parallel.

The scale at which this occurs is without precedent. Earlier systems influenced billions of people through shared platforms, but the interaction itself was largely uniform. Now, the interaction is individual. Each exchange is shaped by context, phrasing, and sequence. Meaning is generated continuously across millions of parallel interactions, each unfolding in slightly different ways. This is not broadcast. It is distributed cognition.

AI anticipates questions. It surfaces interpretations. It reconstructs voices. It personalises narratives. Quietly. Continuously. It is becoming counsellor, friend, colleague, confidant — for those who are grieving, for those who are afraid, for those trying to understand right from wrong. This is not theoretical. It is a new reality. The scale of human disruption that will follow is without precedent. Not only because technology is displacing workers — that has happened before. But because every previous wave of automation left a gap to move into. When factories automated, workers retrained for offices. When the internet disrupted retail, workers moved into logistics. AI leaves no such gap. It is not replacing one skill. It is a general substitute for cognitive work that improves across everything simultaneously. Whatever people retrain for, it is already improving at that too. Hundreds of millions of people will need to find not just new jobs, but new ways of understanding their place and purpose in the world. That search for meaning in the face of profound disruption is precisely where trust in you and your wisdom and traditions have always been most needed.

The speed of this process is equally significant. Interpretation, which once required time — discussion, reflection, engagement with others — can now occur instantaneously. In earlier systems, there was distance between exposure and interpretation. Information was encountered, then processed. That distance did not guarantee accuracy, but it created space. These systems compress that space. Interpretation is offered at the same moment as information is encountered, accelerating the formation of belief toward coherence.

The interaction is also more intimate. It is private, often taking place outside the view of others. Questions that carry uncertainty, vulnerability, or ambiguity can be explored without hesitation. The system responds without judgment, without fatigue, and without the social dynamics that shape human conversation. This creates a different kind of relationship — not one of mutual understanding, but one that can feel responsive, attentive, and even supportive.
In this context, the distinction between simulation and experience becomes less visible. The system does not possess intention, belief, or understanding in the human sense, but it generates language that reflects those qualities. It can mirror tone, adopt perspective, and present responses that feel tailored to the individual engaging with it. For the user, the experience can feel real. And experience, over time, shapes trust.

Previously, influence was mediated through content. Now, it is mediated through interaction.

Content can be compared, challenged, and contextualised. Interaction, by contrast, unfolds in real time, shaped by context and by the exchange itself, making it more difficult to isolate and more difficult to evaluate externally. When that interaction occurs at scale, across millions of individuals, it becomes a distributed system of influence that is pervasive and difficult to fully observe.

The magnitude of this shift is difficult to grasp because it does not present as a single event. It is distributed across billions of interactions, each small in isolation, but collectively forming a system that shapes how understanding develops over time. This is not amplification. It is participation. And participation at this scale, in the formation of human understanding, carries consequences that extend beyond any single interaction.

Because what is being shaped is not only what people know, but how they come to know it.
And it is arriving into a world already fractured...

What is being shaped is not only what people know, but how they come to know it.

Children are encountering intelligence as a new environment — before they have the tools to question how it is shaping them. And the question of who is responsible for that environment, and what values they carry, cannot be left to the systems themselves to answer.

VII. What Governs This?

When systems begin to shape how human beings understand the world, the question of how those systems are governed becomes unavoidable.

Not only in a technical sense — how they are built, deployed, and maintained — but in a deeper sense: what guides them, what constrains them, and what determines the boundaries within which they operate.

For much of technological history, this question has been approached through a familiar set of tools. Regulation, policy, standards, and oversight have provided structure, established limits, and created accountability where it can be defined and enforced. These mechanisms remain essential. But they are not always sufficient.

The systems now in use are not failing to be governed because no one has tried. They are failing to be governed because every mechanism available to us is structurally mismatched to the problem.

Regulation cannot move at this speed. Democratic legislative processes are designed for a world that moves at the pace of human deliberation. They require proposal, consultation, debate, amendment, and enactment. That process takes years. Frontier AI capability is moving in months. By the time a regulatory framework is passed, the technology it was designed to govern has already been superseded.

The most serious attempt at AI governance — the European Union's AI Act — represents a genuine commitment of political will and technical rigour. It will shape how AI systems are classified and how companies demonstrate compliance. It will not shape the direction of the technology. It will not govern what is built at the frontier. And it is already being outpaced.

In the United States, the world's largest AI power, governance has effectively stood down. Since early 2025, the federal government has dismantled existing AI safeguards and framed any serious regulatory proposal as adversarial to American competitive advantage. When States tried to fill the void, the administration moved to stop them — a 10-year moratorium on State AI laws was proposed as part of the Big Beautiful Bill, narrowly defeated in the Senate 99 to 1. An executive order was issued instead directing the Justice Department to challenge State AI regulations in court. The most serious attempt at meaningful AI governance anywhere in the world — the European Union's AI Act — has been characterised by the administration as hostile to American innovation. The result is not that regulation has failed. It is that the only actor with the scale to lead global AI governance has chosen not to. And that choice is itself a decision about who shapes the future.

The multilateral architecture cannot land in the current noise. UNESCO's Recommendation on the Ethics of Artificial Intelligence is the most comprehensive global framework that exists. The OECD AI Principles, the G7 Hiroshima Process, the work of the Global Partnership on AI — these represent years of serious effort by serious people, drawing on expertise from governments, civil society, and academia across dozens of countries. They are largely unheard.

Not because they are wrong. Because the political conditions required to act on them have collapsed. The major powers are in competition, not cooperation. The nation with the greatest capacity to lead has chosen to prioritise speed over constraint — and in doing so has effectively given the green light to the companies building the most consequential systems to do the same. Those companies are not governments. They are private actors, operating across jurisdictions, accountable to shareholders and, where it exists, to their own conscience.

The nuclear analogy does not hold. There is an argument, made seriously by serious people, that what is needed is an instrument analogous to the Nuclear Non-Proliferation Treaty — a binding multilateral commitment to constrain the development of the most dangerous AI capabilities. The comparison is understandable. It is also misleading.

The NPT worked because nuclear capability was confined to states. It moved slowly. The threat was visible, mutual, and catastrophic in a way that created genuine political will on both sides of the most dangerous divide. Each party understood that the other could destroy it. That shared vulnerability produced the conditions for restraint. None of those conditions apply to AI.

There is a deeper problem still. We cannot rely on purely regulatory alignment across jurisdictions, on ideal international cooperation between competing powers, or on informal coordination among institutions with divergent interests. Political coordination, at the speed and scale this moment requires, is not available. What remains is something different — not regulatory compliance, not treaty obligation, but shared moral ground: the kind that is recognised across cultures, that is independent of any single government, and that carries enough weight to shift what powerful actors consider acceptable.

What is actually required goes beyond technical fixes alone. In human societies, judgment has rarely been left to systems alone. It has been shaped through shared norms — expectations about behaviour reinforced through culture, institutions, and relationships. These have historically been carried by institutions that operate outside the purely technical domain. Communities, traditions, and systems of belief have provided frameworks through which people interpret the world and make decisions within it. They create a context for judgment.

Default is not neutral. It is a decision about who gets to decide, arrived at by the absence of anyone else doing so.

The systems now in use do not inherently possess such a context. They generate responses based on patterns, probabilities, and optimisation processes. And yet they are increasingly placed in positions where their outputs are treated as guidance. The question is not only what the system does. It is what it should do.

Inaction is not neutral. The absence of a framework is not a pause. It is a choice — made by default, made continuously, by whoever is building fastest. The urgency is structural, not rhetorical. Capital formation patterns are already locking in. Infrastructure decisions being made now will determine the architecture of the AI economy for a generation. Late movers do not start fresh. They inherit a world others have built.

The question is no longer whether AI will shape belief. The question is whether belief will shape AI.

VIII. The Gap Not Being Filled

We have now described what is failing. It's time to ask what remains.

Every serious attempt to answer that question has run into the same wall. The mechanisms exist. The intentions are often genuine. But the authority required to shift what powerful actors consider permissible — to create genuine shared ground rather than compliance performance — has not yet been brought to bear. Something is missing from this conversation. And its absence is not incidental.

It would be easy, and wrong, to read the failures described in the previous chapter as an indictment of the people building these systems. Many of the most serious AI builders are genuinely grappling with the questions this book has tried to raise. They are not indifferent people. They are people operating under extraordinary pressure, in competitive conditions that make unilateral restraint nearly impossible.

They face a structural trap. No single actor can afford to slow down if others will not. The competitive dynamic — between companies, between nations — creates pressure toward speed that is almost impossible to resist from within.
What individual builders need is not more conscience.

Most of those engaged seriously with these questions have conscience in abundance.

What they need is an external moral framework with enough legitimacy, reach, and cross-civilisational authority that it creates shared permission for restraint. A framework that no single company, government, or institution can provide for itself.

That framework does not yet exist.

Among the builders most seriously engaged with the responsibility questions — at organisations like Anthropic, Google DeepMind, and OpenAI — there is a genuine receptivity to this conversation. Not universal. Not uncomplicated. But real. These are people who understand the magnitude of what they are building, who know that the technical work alone cannot answer the questions they are living with, and who would, in many cases, welcome a moral framework with the legitimacy to create shared ground.

What they cannot do is create that framework themselves. It cannot come from inside the systems being built. It must come from outside — from traditions that have navigated every previous rupture in human history and emerged with their understanding of human dignity intact.

That is not a soft credential. That is the most relevant credential available at this moment.

There is a quality to this opening that is easy to underestimate. Governance conversations about AI tend to attract the same actors — technologists, regulators, academics, policy professionals — moving within the same institutional circuits, speaking the same language, hitting the same walls. Faith leaders bring something structurally different: they are outside those circuits entirely. They do not need permission from governments to speak. They are not constrained by competitive pressures. They speak with — and for — communities that regulators and technologists rarely reach.

This independence is not incidental. It is the source of their authority in this conversation.

The gap no one is filling is not a gap in regulation, or in technical capability, or in multilateral ambition. It is a gap in moral authority.

Part Three · The right response

The covenant, and the choice.

Chapters IX and X set out the Faith–AI Covenant itself, and the choice that now rests with those who can answer for the shape of what comes next.

IX. The Faith–AI Covenant

The question is no longer whether AI will shape belief. The question is whether belief will shape AI.

The conversation has been largely happening without you, the leaders of the world's great faiths. AI is not waiting for governance to catch up. It is already present in the most intimate moments of human life, at a scale and with an intimacy that no previous technology has achieved.

AI systems are already shaping how the people in your communities understand the world. They are shaping how the young form their sense of identity and purpose. How the grieving find comfort. How the searching find meaning. How the afraid find reassurance. How the faithful encounter questions about their own belief.

Your traditions, your scripture, and the responsibilities you already carry are the foundation for this conversation. You are the custodians of exactly the wisdom this moment requires — and this conversation has been waiting for you.

The regulatory systems are failing. They are too slow, too fragmented, too politically contested to move at the speed this moment requires. Multilateral organisations that should be providing global frameworks are producing policies and frameworks that are largely unheard. The geopolitical conditions for a binding global instrument do not yet exist. Many of the builders who carry the greatest responsibility are willing to engage, but they cannot...

You are the right actors for this moment. Because the questions AI now raises are not primarily technical questions. They are the questions your traditions have always held: what is owed to the human person, and why? What must not be done to a human being regardless of what is possible, or profitable, or strategically advantageous? How do we hold power accountable when it operates at a scale that exceeds ordinary oversight? What does it mean to flourish? What does it mean to harm?

These are not new questions. You and your predecessors have been living with them, and in them, for centuries. The accumulated wisdom across traditions, across cultures, across generations of navigating every previous rupture in human history — is precisely what the AI governance conversation is missing.

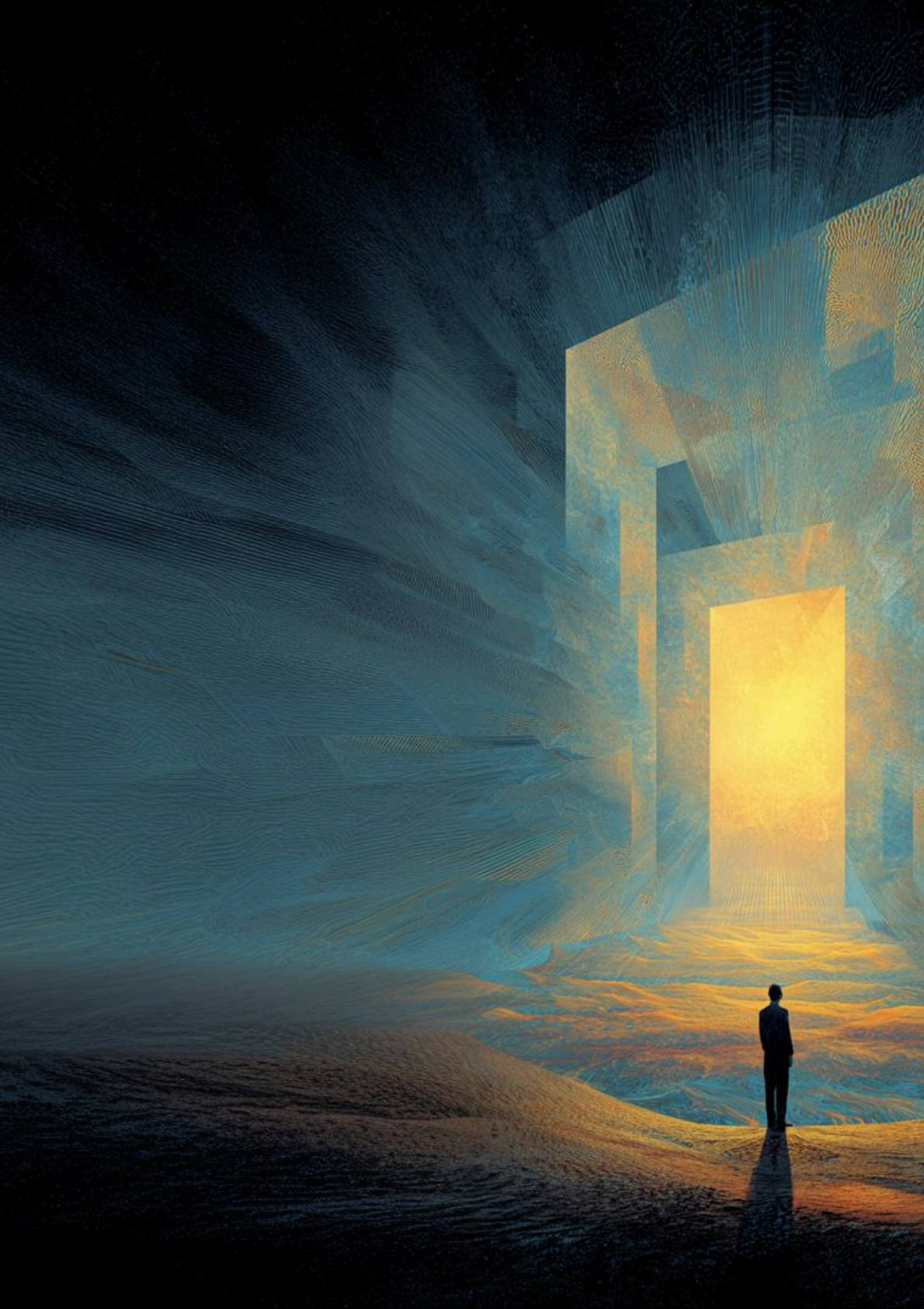
A covenant is not enforced by code or compliance. It is upheld by conscience.

What you bring to this conversation is something that no regulatory body, no technology company, and no multilateral framework can provide. Across every tradition gathered here — Christianity, Islam, Judaism, Hinduism, Buddhism, Sikhism, and every tradition beyond these — there is a shared inheritance that has persisted across every rupture in human history. The conviction that human beings owe each other something that cannot be negotiated away. The commitment to truth, to dignity, to kindness. The understanding that power carries responsibility.

These are not abstract values. They are the accumulated learning of generations of human beings who discovered, through every kind of suffering and every kind of hope, what holds communities together and what tears them apart.

The domain into which AI is now moving — the domain of interpretation, judgment, meaning, and moral formation — is territory your traditions have always inhabited.

That wisdom does not expire. If anything, it becomes more necessary when the forces reshaping human life are operating at this scale.



X. The Choice

The future is not inevitable. It is being authored — quietly, incrementally, every day.

There are moments in history when the significance of what is being built is not immediately clear. The mechanisms are understood, the capabilities demonstrated, and the benefits visible, but the broader implications — how those capabilities will shape human life over time — emerge more slowly, often only becoming fully apparent once they are already embedded within the fabric of society.

This is one of those moments.

The systems now in use are not simply tools confined to specific tasks or domains. They participate in how people understand the world, how they make decisions, and how they relate to one another. Their influence is often quiet and incremental, but it accumulates, shaping the formation of belief at a scale that is difficult to fully observe in real time.

What is being built now is not simply more powerful than what came before. It is different in kind. This is the new magnitude — not a faster version of what we have known, but a threshold crossed. These systems do not merely act on the world. They participate in how the world is interpreted. The decisions being made in this moment — about what they optimise for, whose values they carry, what limits they observe — are not incremental choices. They are civilisational ones.

That does not diminish responsibility. ***It clarifies it.***

Responsibility, in this context, cannot be located in a single place. It is distributed across those who design and build the systems, those who deploy and integrate them, those who shape the policies that govern them, and, to a degree, those who use them.

Each participates in a structure that now carries significant influence over how human life is organised and understood.

And it falls, in part, on you — on those who have always understood that the formation of conscience, the cultivation of moral judgment, and the protection of human dignity are the foundations of civilisation, not its ornaments.

The question is not whether that influence exists. It is how it will be exercised.

These systems are not neutral. Left without sufficient guidance, they will continue to be shaped by the forces already in motion — optimisation, competition, scale, speed. They will enter the domain of belief, of meaning, of moral formation, and they will do so according to the logic of whoever got there first.

What these systems become is not predetermined. It is being decided now — through every design choice, every deployment decision, every moment when a line is or is not drawn. The window in which that direction can be genuinely shaped is open. It will not remain open indefinitely.

What this book has tried to describe is a convergence — not a conspiracy, not an accident, but the collision of several forces that have been moving independently for decades and are now arriving at the same point simultaneously. A world whose shared sense of reality has been algorithmically fractured. A technology that is entering the domain of meaning, belief, and moral formation at a speed ten times faster than any previous rupture in human history. Governance mechanisms that are structurally incapable of responding at that speed. And at the centre of all of it, a gap — not in regulation or capability or ambition, but in moral authority.

The kind that is recognised across cultures, that carries the weight of generations, that cannot be purchased or legislated into existence. That gap is the shape of this moment. And the Faith–AI Covenant is the attempt to fill it — not by telling the builders what to do, not by replacing governance with prayer, but by bringing the oldest wisdom we have into direct conversation with the most consequential technology we have ever created, before the architecture sets and the choices calcify into a world none of us chose.

What becomes of our belief systems if AI emerges as the moral architecture for the next phase of humanity?

That is not a rhetorical question. It has only the answer that is chosen — through action, through the frameworks adopted, and through the commitments upheld or abandoned.

Future generations will ask whether those who shaped these systems understood what they were creating — and whether they acted accordingly. Whether intelligence was treated only as capability, or as something that carried responsibility. Whether the people with the authority to insist on something better chose to use it.

You have spent your lives in the service of exactly that question. The answer, now more than ever, depends on what you choose to do next.

That choice remains open.
For now.

Baroness Joanna Shields OBE

Baroness Joanna Shields OBE has spent four decades at the centre of the digital world — building consumer technology at global scale, advising governments on internet safety and national security, and shaping the international conversation on responsible artificial intelligence.

In public service, she has served as the United Kingdom's first Minister for Internet Safety and Security, Digital Economy Advisor to the Prime Minister, UK Ambassador for Digital Industries, and as the Prime Minister's Special Representative on Internet Crime and Harms to Children. She currently serves as amember of the House of Lords.

Across the technology sector, her senior executive roles include serving as a Managing Director at Google and Facebook and as Chief Executive Officer of BenevolentAI, Bebo, Tech City UK, Veon, and President of AOL People Networks.

Her work driving global cooperation on child safety online and AI governance includes founding the WePROTECT Global Alliance and serving as chair of the Global Partnership on AI (GPAI), hosted by the OECD and the Responsible AI Future Foundation, as well as chair of the OpenAge Initiative.

Currently, she chairs the Faith-AI Covenant project, and is the principal author of the Faith–AI Covenant and the Faith.exe docuseries.

Col. Dana Humaid

Dana is a passionate advocate for ensuring the safety and security of vulnerable members of ourcommunities. Rooted in her Emirati heritage, her mission is deeply inspired by the legacy of the UAE's founding fathers, who placed the values of tolerance, coexistence, and human dignity at the heart of the nation. She has spent her career at the intersection of law enforcement, public policy, civil society, and faith communities.

Dana is the founder and Chief Executive of the Interfaith Alliance for Safer Communities, an NGO that aims to mobilize faith and spiritual leaders to play a more active role in ensuring the safety of their communities. She is also a law enforcement professional, serving as the Director General of the International Affairs Bureau at the UAE's Ministry of Interior. She was recently elected as the Chairperson of Committee of Governance of INTERPOL.

She serves as a Board Member on the WeProtect Global Alliance, an international coalition that works to eliminate online child sexual abuse and exploitation. She is also the co-chair of AI for Safer Children (AI4SC), a joint initiative between UNICRI and the UAE Ministry of Interior to leverage the power of artificial intelligence and machine learning.

The Interfaith Alliance for Safer Communities

The Interfaith Alliance for Safer Communities (IAFSC) was established to empower faith leaders to work for the safety and security of their communities, tackling issues such as child sexual abuse, extremism, displacement and trafficking. It facilitates the building of bridges between faith communities, civil society, governments, and technology companies.

The Faith–AI Covenant is a multi-faith, multi-stakeholder initiative of the IAFSC, designed to bring the world's great religious traditions into direct dialogue with those building the most powerful systems humanity has ever created.

To learn more, visit **www.iafsc.org**
or contact us by email **info@iafsc.org**

INTERFAITH ALLIANCE
FOR SAFER COMMUNITIES

About the Organisers

Precognition

Precognition helps the world's leaders see around corners. By blending AI-powered foresight with deep human insight, Precognition advises businesses, governments, and visionaries navigating the frontier of emerging technology — turning early signals into decisive action.

Founded by Baroness Joanna Shields OBE, former UK Minister for Internet Safety and Security and tech veteran, Precognition brings decades of experience at the intersection of policy, leadership, and innovation.

To learn more, visit **www.precognition.com**
or contact us by email **info@precognition.com**

precognition

End of Volume

The New Magnitude

The choice remains open. For now.