

A guide to
**FINDING YOUR PLACE IN
THE BESCI & AI
LANDSCAPE**

Elina Halonen - June 2026

v. 1.0

Introduction

If you are curious about the relationship between behavioural science and AI, you may already have noticed that the field often seems difficult to pin down. Different people use the same language to describe very different kinds of work, and many conversations assume a level of familiarity that newcomers may not yet have. That is not simply confusion. The space is still forming, with researchers, practitioners, designers, policy people, and others approaching different problems, at different levels, and with different ideas about what behavioural science might be for.

That can make the field difficult to enter, even for people who can see that AI raises deeply behavioural questions. You may recognise that your existing expertise is relevant, but still struggle to work out where it fits, what kind of contribution you could make, or how to explain that contribution to others. The door may be there, but it is not always helpfully labelled.

Part of the difficulty in navigating BeSci-AI is that people in the space are often doing fundamentally different kinds of work: trying to understand AI's effects, shape AI systems, implement AI in organisations, govern AI's consequences, or question the assumptions behind how AI problems are framed. These are all important, but they are not the same kind of work, and they do not call for the same forms of expertise.

I have written this as the guide I wish had existed when I first began trying to understand where behavioural science might fit in relation to AI. It took time to see how many different kinds of work were being gathered under similar language, and how many possible routes there were into the space. This guide is not offered as a final map of the field, but as a working orientation tool for others who may be trying to find their way into it more quickly.

This guide draws on a review of the academic literature to name seven orientations that recur across the space. Those academic roots are useful because they show where different ways of thinking have come from, but the guide is written for researchers, practitioners, designers, strategists, policy people, and others who can see that behavioural science and AI belong in conversation, but who may not yet know where their own experience, interests, or responsibilities fit.

The goal is practical: to give you language for your own position, and a way to understand why someone else working in BeSci-AI might mean something quite different by it. It is also an invitation to get involved from a clearer place. Behavioural science will be needed in many different forms as AI reshapes work, institutions, services, and everyday decision-making, and many people who have something useful to contribute may not yet recognise where their route into the work begins.

To get the most from the guide, read the seven orientations and ask: which feels most like my default? Which do I tend to overlook? Which would most improve my thinking if I took it seriously?

Where the [*Field Guide to Problem Orientations in Behavioural Science and AI*](#) mapped the literature, this guide turns that map toward professional practice. It is designed to help people working across behavioural science and AI locate their own default orientation, understand why collaboration across the space can feel oddly misaligned, and identify where their way of thinking may be recognised in organisations, research settings, or advisory work.

How to use this guide

One way to start locating yourself in the BeSci-AI landscape is to look at the level where your attention naturally goes first. These levels are connected, but they are not interchangeable:

- **Individual:** how AI shapes judgement, trust, behaviour, motivation, attention, agency, reliance, cognitive offloading, and emotional responses
- **Interactional:** what emerges through repeated human-AI encounters, including adaptation, relational dynamics, and mutual influence over time
- **Organisational:** workflows, routines, incentives, adoption barriers, capability gaps, and institutional embedding
- **Systemic:** governance, power, public legitimacy, markets, and wider social consequences.

This distinction between levels matters because a problem may appear at one level while being caused, sustained, or solved at another. For example:

- A decline in user trust might appear at the individual level, but stem from poor governance at the systemic level.
- Resistance to AI adoption in organisations might appear at the organisational level, but reflect deeper concerns about fairness and accountability at the systemic level.
- An AI system that performs well in testing might appear successful at the interactional level, but still fail when it meets real organisational incentives and workflows.

The seven orientations in this guide are built from these differences in level of analysis. Each orientation combines a preferred level of analysis with a particular object of attention and a particular idea of what behavioural science is for:

- a user-centred orientation usually starts from individual and interactional effects
- a deployment-centred orientation starts from organisational routines and constraints
- a governance-centred orientation starts from institutional and systemic consequences

These are meant to be seen as recurring patterns in how people define the problem, rather than as rigid boxes. This is why two people can both say they work on behavioural science and AI while meaning quite different things. One may be asking how AI affects trust and judgement, while another may be asking how AI can be embedded into organisational workflows, or whether the way AI is being used is fair, transparent, and acceptable to the public. They are working from different starting points, which means they may notice different problems, accept different kinds of evidence, and disagree about what would count as a useful answer.

You can use the guide in three ways:

- **As a self-location tool:** helping you identify which orientation most closely matches your instinctive way of understanding behavioural science and AI problems.
- **As a collaboration tool:** helping you anticipate which other orientations are likely to work smoothly with yours, and where misunderstandings or tensions may arise because people are approaching the problem differently.
- **As a positioning tool:** helping you recognise where this way of thinking is likely to be valued, supported, hired for, or potentially misunderstood in professional and institutional settings.

Taken together, these uses should help you read the landscape less as a single field with one shared agenda, and more as a set of overlapping professional and intellectual positions. The next section describes the seven orientations in more detail, so you can begin to locate your own default stance and recognise where others may be working from a different one.

The seven orientations

The following profiles describe each orientation in terms of the core question it revolves around, what tends to draw attention, and how that orientation tends to understand AI and the role of behavioural science within it. The wording is deliberately suggestive rather than diagnostic. Most people will recognise elements of more than one orientation, but one or two may feel more like their default starting point.

User-centred

Core question you're interested in: What is AI doing to people's judgement, behaviour, trust, attention, motivation, and sense of agency?

Focus of attention: Reliance, over-reliance, persuasion, cognitive offloading, emotional response, adaptation, and how behaviour changes through repeated interaction with AI.

How you probably see AI: AI is something people encounter, adapt to, and are shaped by over time. You are interested in AI as an experience rather than a system: how it feels to use, how it changes what people do, and how repeated interaction accumulates into something more than a series of individual decisions.

The role of behavioural science for you: Behavioural science gives you the conceptual and empirical toolkit for making human responses making human responses visible and easier to describe: frameworks for reliance, trust calibration, cognitive offloading, and behavioural adaptation - the vocabulary for naming what is happening to people and why it matters.

Design-centred

Core question you're interested in: How can behavioural science help design AI systems and interfaces that actually fit how people think and decide?

Focus of attention: Usability, cognitive fit, decision support, interface structure, friction, feedback, and choice architecture.

How you probably see AI: AI is something to be shaped upstream. The system is not yet fixed, and the design decisions being made now will determine whether people can actually use it well. You are interested in AI as an artefact - something that can be built better or worse, and where behavioural science has something specific to contribute before problems become embedded.

The role of behavioural science for you: Behavioural science provides design principles: knowledge about cognitive fit, friction, feedback, and choice architecture that translates how people actually think into concrete design specifications, rather than discovering after deployment that the system works against its users.

Deployment-centred

Core question you're interested in: What does it actually take to make AI work in real organisations with real workflows, incentives, and constraints?

Focus of attention: Embedding, stakeholder buy-in, implementation barriers, capability gaps, behavioural bottlenecks inside organisations, and operational friction.

How you probably see AI: AI is something that has to survive contact with practice. The gap between a good idea and a working implementation is where most projects actually live, and that gap is primarily behavioural and organisational rather than technical. You are less interested in whether something is well-designed in principle than in whether it will hold up when real teams and competing priorities get involved. This differs from intervention-centred work because the primary concern is embedding AI into organisational practice, rather than using AI to tailor or scale behavioural interventions themselves.

The role of behavioural science for you: Behavioural science gives you a toolkit for understanding implementation as a behavioural problem — identifying where routines resist change, where incentives undermine adoption, and where stakeholder behaviour will determine whether a system embeds or quietly gets abandoned.

Agent-analytic

Core question you're interested in: Can you study AI systems the way you would study behaviour by observing patterns, tendencies, and interaction effects rather than just inspecting the code?

Focus of attention: Consistency, adaptation, emergent behaviour, multi-agent dynamics, behavioural signatures, and interaction patterns that only become visible over time.

How you probably see AI: AI is an object of behavioural inquiry, not just a tool or a policy problem. You are interested in what AI systems actually do when they interact with people and environments over time: how they adapt, what patterns emerge, what their behavioural signatures look like across contexts.

The role of behavioural science for you: Behavioural science provides the conceptual vocabulary and observational methods for studying AI systems as behaving entities: frameworks for consistency, adaptation, and interaction effects applied to systems rather than people, or to the dynamics that emerge when both are in the room.

Intervention-centred

Core question you're interested in: How can AI help behavioural science interventions reach more people, more precisely, and with better results?

Focus of attention: Personalisation, prediction, adaptive delivery, targeting, experimentation, monitoring, and scalable intervention design.

How you probably see AI: AI is an amplifier and enabler. It makes it possible to personalise interventions at scale, run continuous experiments, and adapt delivery in real time. You are interested in AI as infrastructure that makes behavioural science more powerful and more useful in practice. This differs from deployment-centred work because the primary concern is the intervention logic and its delivery, rather than the organisational embedding of AI as a workplace or service system.

The role of behavioural science for you: Behavioural science provides the intervention logic - the theory of change, the specification of target behaviours, the design of mechanisms, and the criteria for success. AI executes, adapts, and scales what behavioural science has designed.

Governance-centred

Core question you're interested in: How do we ensure that an understanding of human behaviour is reflected in the rules, responsibilities, and oversight to make sure AI is used fairly, accountably, and in the public interest?

Focus of attention: Accountability, oversight, public consequences, distribution of harms and benefits, institutional fit, and whether AI arrangements can withstand democratic scrutiny.

How you probably see AI: AI is a sociotechnical system with consequences that extend far beyond the immediate interaction. You are less interested in what a system does in a controlled setting than in how it distributes benefits and harms, whether anyone is accountable when things go wrong, and whether the arrangements around it can withstand public scrutiny.

The role of behavioural science for you: Behavioural science helps translate abstract governance principles into something grounded in how people actually behave - how they respond to oversight, what makes accountability feel real rather than performative, and how behavioural effects should inform regulatory design.

Frame-critical

Core question you're interested in: What assumptions are built into the way BeSci-AI is being framed and what is that framing doing, for whom, and at whose expense?

Focus of attention: Mismatched categories, inflated claims, conceptual stretching, hidden assumptions, misplaced certainty, and value choices hidden in apparently neutral language

How you probably see AI: AI is partly a rhetorical object - something being described and defined in ways that serve particular interests and foreclose particular questions. You are alert to how framing shapes what gets counted as a problem and what responses seem natural. The name "behavioural science and AI" is itself something to interrogate rather than accept as given.

The role of behavioural science for you: Behavioural science functions as a critical lens. It helps you identify motivated reasoning, conceptual stretching, inflated claims, and cases where political or value-based choices are made to appear technical, neutral, or inevitable. You use it to ask what work a framing is doing before accepting the problem as stated.

Summary of the seven orientations

The table below is meant as a quick reference that makes the differences easier to scan, but the orientations are not meant to function as fixed professional identities. The value of the table is that it makes the starting point visible: what kind of problem someone notices first, what they treat as evidence, and what they are likely to think needs attention next.

Orientation	Core question	Focus of attention
User-centred	What is AI doing to people's judgement, behaviour, trust, attention, motivation, or sense of agency?	Reliance, over-reliance, anthropomorphism, persuasion, adaptation, cognitive offloading, emotional response, behaviour change through repeated interaction.
Design-centred	How can behavioural science help design better AI systems, interfaces, and interactions?	Human factors, usability, cognitive fit, decision support, interface structure, friction, feedback, prompt and choice architecture.
Deployment-centred	Why do apparently sensible BeSci-AI ideas fail when they meet organisations, routines, workflows, and incentives?	Embedding, capability, stakeholder buy-in, implementation barriers, behavioural bottlenecks inside organisations, operational frictions.
Agent-analytic	Can AI systems themselves be studied as behaving entities, with patterns, tendencies, strategies, and interaction effects?	Consistency, adaptation, emergent behaviour, multi-agent dynamics, behavioural signatures, interaction patterns that are not reducible to code inspection.
Intervention-centred	How can AI be used to deliver, tailor, scale, or optimise behavioural science interventions?	Personalisation, prediction, experimentation, adaptive delivery, targeting, monitoring, scalable intervention design.
Governance-centred	What kinds of institutions, rules, responsibilities, and public settlements are needed around AI?	Legitimacy, accountability, oversight, public consequences, distribution of harms and benefits, institutional fit, democratic acceptability.
Frame-critical	What assumptions are built into this framing, and what work is that framing quietly doing?	Category mistakes, rhetorical inflation, conceptual stretching, depoliticisation, hidden ideology, misplaced certainty, intellectual overreach.

What each orientation offers

Once you have a sense of where your attention usually goes first, the next question is what that starting point helps you see and what it may make easier to miss. Each orientation brings a particular kind of value, but focusing attention in one place inevitably creates blind spots elsewhere.

Looking at both sides can help you use your own orientation more deliberately, and recognise when you may need another perspective, either by involving someone who works from a different orientation or by learning enough about that area to spot risks, question assumptions, understand consequences, or make an idea workable in practice.

User-centred

Your strength: You keep people's actual experience in view when others have moved on to systems, metrics, or abstractions. You notice effects that technical performance measures may miss entirely: the quiet erosion of confidence, the gradual redistribution of cognitive effort, the way trust forms and breaks through repeated use. You insist that what AI does to people counts as evidence.

What you risk missing: The individual-level story can become the whole story, even when the deeper issue lies in institutional design, incentives, or governance. For example, people may appear to mistrust an AI system because they are anxious, resistant, or poorly informed, when the more important problem is that the system is being introduced without clear accountability, realistic support, or a credible way to challenge its outputs. You may document human effects clearly while the conditions producing them remain underexamined.

Worth borrowing from or teaming up with:

- Governance-centred thinking can help you ask who is responsible when things go wrong, not only how people experience the system. It brings attention to questions about accountability, oversight, incentives, and institutional responsibility that may sit behind the effects you are observing.
- Deployment-centred thinking can help you understand why the effects you observe keep happening in practice. It draws attention to workflows, organisational pressures, capability gaps, and operational realities that shape how people actually end up using a system.

Design-centred

Your strength: You make the gap between intended and actual use easier to address before the system is built or embedded. You translate knowledge about how people think, decide, notice risk, recover from mistakes, or become overloaded into concrete design choices. This orientation is often strongest when systems are still taking shape, because it tries to prevent predictable behavioural problems from becoming embedded into the product, interface, or workflow.

What you risk missing: You can assume that the central problem is poor design when the real disagreement is about goals, power, responsibility, or public legitimacy. For example, a system may fail because people fundamentally disagree with how it is being used, not because the interface is confusing or the workflow is poorly designed. A system can be easy to use and still be the wrong system to build or deploy.

Worth borrowing from or teaming up with:

- User-centred thinking can help you keep the design grounded in what people actually experience, not only what the system is intended to do. It brings attention to whether people understand the system, trust it appropriately, recover from mistakes, and experience the interaction as useful rather than merely well designed.
- Governance-centred thinking can help you examine whether a well-designed system is also fair, accountable, and publicly defensible. It brings in questions about oversight, responsibility, and who carries the consequences if the system causes harm or fails in practice.

Deployment-centred

Your strength: You bring practical realism to a field that can otherwise stay too comfortable in concepts, principles, or prototypes. You understand where promising ideas tend to fail: in workflows, routines, incentives, procurement, capability gaps, and everyday organisational pressures. You take seriously the fact that resistance to a system is often rational from the perspective of the people resisting it, because they may be carrying risks, extra work, or responsibility that the designers have underestimated.

What you may miss: The focus on making systems work can subtly turn deeper disagreements into adoption problems. For example, staff reluctance to use a system may be interpreted as resistance to change, when they may actually be worried that the system creates unrealistic workloads, weakens professional judgement, shifts accountability onto frontline staff, or produces unfair outcomes for the people affected by it. In these situations, the problem may not be that people are refusing to adapt. It may be that they have recognised a genuine risk or trade-off that has not been taken seriously enough.

Worth borrowing from or teaming up with:

- User-centred thinking can help you check whether a system that works operationally is also working well for the people using it. A process may run smoothly while still creating confusion, dependence, frustration, or loss of confidence for users, staff, or affected communities.
- Governance-centred thinking can help you check whether adoption problems are symptoms of a wider accountability problem. It draws attention to whether the system has clear responsibility, credible oversight, and institutional arrangements that people can trust.

Agent-analytic

Your strength: You open up a layer of analysis that neither technical inspection nor user research usually reaches. You look for patterns that only appear once AI systems are used repeatedly across different contexts: recurring behaviours, shifts in outputs, adaptation through interaction, or unexpected dynamics between systems and people. This orientation is useful when the important question is not only whether a system passed a benchmark or whether users liked it, but what the system actually does over time once it enters a messy human environment.

What you may miss: It is easy to become absorbed in the behaviour of the system and lose sight of why those patterns matter. For example, a system may show interesting variation across contexts, but that variation only becomes important once it affects users, organisations, accountability, or risk. There is also a real risk that behavioural language applied to AI systems can become too human-like, as if the system has intentions, motives, or understanding, when what you may actually be observing is a pattern in outputs, responses, or interaction.

Worth borrowing from or teaming up with:

- User-centred thinking can help you connect system patterns back to human experience. It asks what those patterns mean for people's trust, judgement, confidence, dependence, or ability to challenge the system.
- Frame-critical thinking can help you notice when language about AI behaviour is becoming too human-like or too confident. It can also help you ask whether words such as "agent," "intention," or "understanding" are clarifying what the system does, or making it seem more human than it is.

Intervention-centred

Your strength: You connect evidence with scale. You take seriously the practical challenge of getting interventions to more people, more precisely, and with stronger effects. You stay close to measurable outcomes, experimentation, and real-world impact. This orientation is often strongest when the question is how to move from promising behavioural evidence to something that can work consistently across large populations, services, or systems.

What you may miss: You can move too quickly from understanding behaviour to optimising it. For example, an intervention may improve compliance, increase uptake, or change behaviour in measurable ways while still raising difficult questions about manipulation, fairness, autonomy, or unequal impact across groups. The question of whether an intervention should be scaled, including whether its values and trade-offs remain acceptable at population level, may receive less attention than the question of whether it can be scaled.

Worth borrowing from or teaming up with:

- Governance-centred thinking can help you ask who benefits, who carries the risk, and who has authority to shape behaviour at scale. It brings in questions about legitimacy, accountability, oversight, and whether large-scale behavioural influence can be publicly explained and justified.
- Frame-critical thinking can help you ask what assumptions sit behind the intervention before you focus on optimising it. It draws attention to who defines the desired outcome, what kinds of behaviour are being treated as problems, and whether the intervention is solving a genuine issue or mainly making an existing system more efficient.

Governance-centred

Your strength: You keep responsibility and public consequences in view when other orientations are pulled toward the local and immediate. You ask who gets to define what counts as a good outcome, who is affected, who is accountable, and whether an AI arrangement can be explained and defended beyond the team that built it. This orientation becomes especially important once AI systems begin affecting access to services, professional judgement, public trust, or the distribution of risks and benefits across society.

What you may miss: It is often easier to diagnose the need for governance than to specify how it should work in practice. At its weakest, this orientation produces principles without mechanisms: statements about oversight, accountability, fairness, or transparency that do not yet explain what people are supposed to notice, decide, escalate, record, challenge, or actually do in everyday practice. Governance can sound reassuring at the level of principle while remaining vague operationally.

Worth borrowing from or teaming up with:

- User-centred thinking can help you bring in what people are actually experiencing. It draws attention to whether systems feel challengeable, understandable, trustworthy, or fair from the perspective of the people affected by them.
- Deployment-centred thinking can help you ask how your principles would work on a Tuesday afternoon in a real organisation. It brings attention to the practical realities that determine whether governance arrangements survive contact with workloads, routines, incentives, procurement pressures, and institutional constraints.

Frame-critical

Your strength: You surface assumptions that other orientations may take for granted. You notice when different problems are being treated as if they were the same, when claims sound larger or more settled than the evidence supports, and when apparently neutral language hides values, interests, or political choices. This orientation is especially useful in a rapidly developing field where concepts, labels, and problem definitions are still unstable and where broad terms such as “human-centred,” “responsible,” or “aligned” can mean very different things in practice.

What you may miss: Diagnosis is easier than specification. At its weakest, this orientation becomes much better at identifying what is wrong with a framing than at helping decide what should be built, changed, governed, or tested next. It can also become detached from the practical constraints of implementation, leaving other people to work out how to act once the critique has been made.

Worth borrowing from or teaming up with:

- Design-centred thinking can push you toward saying what should exist, not only what is wrong. It encourages movement from critique into concrete alternatives, workflows, interfaces, or institutional arrangements.
- Intervention-centred thinking can keep asking what should actually happen next. It brings attention back to action, implementation, measurable effects, and whether a proposed alternative can work outside theory or critique.

Orientation, likely homes, and likely roles

Some orientations are easier for organisations to recognise than others:

- Design-centred, deployment-centred, and intervention-centred roles are often easier to hire for because they sound actionable.
- Governance-centred roles are becoming easier to recognise because regulation, assurance, and responsible AI work are growing.
- Frame-critical roles are often harder to place inside organisations unless they sit within academia, strategy, policy, ethics, or public-interest work.
- User-centred roles can be recognised, but sometimes get flattened into generic UX.
- Agent-analytic roles are gaining status in frontier AI contexts, but may be harder to place elsewhere.

Finding your place in this landscape is not only about knowing what kind of work you do. It is also about recognising where that work is likely to be understood. Some orientations translate easily into job titles, briefs, or team functions, while others may need more explanation because their value is indirect, preventive, or strategic rather than immediately operational.

Orientation	Institutional homes	Roles	Objective
User-centred	UX research teams, consumer research, digital product teams, public services, health tech, education tech, trust and safety teams	UX researcher, behavioural scientist, user researcher, service designer, human factors researcher, product insights lead	Understand how AI affects judgement, behaviour, trust, and lived experience
Design-centred	Product design teams, AI labs, HCI groups, innovation teams, design consultancies, responsible design teams	Product or conversation designer, HCI researcher, design strategist, human-centred AI specialist	Improve system design so it better fits human cognition and behaviour
Deployment-centred	Operations teams, transformation programmes, public sector delivery units, enterprise AI teams, implementation consultancies, change teams	Implementation lead, behavioural implementation specialist, change manager, service operations lead, AI adoption lead, organisational design lead	Make AI actually work in workflows, routines, and institutions
Agent-analytic	AI safety groups, evaluation labs, frontier model teams, computational social science groups, research institutes, agentic systems teams	AI evaluator, research scientist, computational social scientist, model behaviour analyst, safety researcher, red team researcher	Study AI systems as exhibiting patterns, tendencies, and interaction effects
Intervention-centred	Health innovation teams, digital therapeutics, public health units, edtech, fintech, policy innovation teams, behavioural intervention consultancies	Behavioural data scientist, intervention designer, digital behaviour change specialist, personalisation lead, experimentation lead, applied scientist	Use AI to tailor, scale, or optimise behavioural interventions
Governance-centred	Regulators, standards bodies, policy units, responsible AI teams, public-interest organisations, legal-policy groups, audit and assurance functions	AI policy adviser, governance lead, responsible AI manager, regulator, standards specialist, ethics and assurance lead	Build legitimate institutional arrangements, oversight, and accountability
Frame-critical	Academia, think-tanks, policy institutes, ethics centres, critical research groups, strategy teams, public-interest organisations	Researcher, strategist, ethicist, critical policy analyst, sociotechnical researcher, senior adviser	Scrutinise assumptions, rhetoric, conceptual overreach, and depoliticising frames

Potential career archetypes

One way to make the orientations more concrete is to ask what kinds of professional roles they tend to produce. The archetypes below are not job titles or career prescriptions - they are shorthand *descriptions of the kind of contribution someone might make when a particular orientation becomes central to their work.*

Most real roles will combine several of these. The value of the archetypes is that they make the underlying work easier to name, especially when formal job titles are vague, inflated, or still catching up with what people are actually being asked to do.

- The product interpreter:** Usually user-centred, or user-centred plus design-centred. Turns messy behavioural effects into design implications. This person notices how AI changes user confidence, reliance, attention, trust, or decision-making, then translates those effects into product choices, interface decisions, service improvements, or research priorities.

- **The organisational translator:** Usually deployment-centred, or design-centred plus deployment-centred. Bridges the gap between concept and institutional routine. This person understands why a promising AI idea may fail once it meets workflows, incentives, capability gaps, internal politics, procurement constraints, and the everyday reality of teams trying to get through the week.
- **The behavioural optimiser:** Usually intervention-centred. Focused on targeting, personalisation, experimentation, and measurable lift. This person is interested in how AI can make behavioural interventions more precise, adaptive, and scalable, while ideally keeping sight of whether the intervention deserves to be scaled in the first place.
- **The model ethnographer:** Usually agent-analytic, or agent-analytic plus user-centred. Interested in what systems do in interaction, not just in benchmark performance. This person studies patterns that emerge over time: how AI systems behave across contexts, how people adapt to them, and what becomes visible only through repeated use.
- **The public legitimacy architect:** Usually governance-centred, or governance-centred plus user-centred. Connects oversight to actual social consequences. This person asks whether rules, responsibilities, review processes, and accountability structures are grounded in how AI is actually used and how its effects are distributed.
- **The frame challenger:** Usually frame-critical, sometimes paired with governance-centred. Spots category mistakes, inflated claims, and respectable language being used to make weak or contested ideas look more solid than they are. This person asks what assumptions are built into the way a problem has been described, what choices are being made to look technical or inevitable, and what becomes harder to question once the frame is accepted.

Note that orientations can be situational as well as personal: someone may work from a governance-centred stance in one project and a deployment-centred stance in another, depending on the problem, institutional setting, and responsibilities involved.

Therefore, these archetypes are best used as rough orientation markers - they are meant to help clarify the kind of work someone is doing beneath the role title, especially in a field where one person's "AI strategy" may mean product design, implementation, governance, optimisation, critique, or all of them squeezed into one heroic but under-scoped job description.

Expanding your range as a professional

Because BeSci-AI projects often involve several different kinds of problem at once, practitioners in this space are likely to benefit from fluency in more than one orientation. The aim is not to become equally competent in all seven, but to know which second orientation might most usefully extend your default one.

Many real projects require people to move between levels. A project may begin with lived experience, but soon raise questions about design, implementation, oversight, or accountability. Another may begin as a governance problem, but still depend on whether people can understand, challenge, or use the system in practice.

Some combinations may prove especially useful because they connect levels that are often separated in practice:

- **User-centred + governance-centred:** useful when decisions about AI need to reflect how people actually experience and respond to these systems, rather than relying only on high-level principles.
- **User-centred + design-centred:** useful when insights about how AI affects trust, attention, judgement, confidence, or cognitive effort need to shape the design of products, interfaces, services, or workflows.
- **Design-centred + deployment-centred:** useful when an AI system needs to work well for users while also fitting into real organisational routines, incentives, and constraints.

- **Deployment-centred + governance-centred:** useful when AI systems need to be introduced in ways that are both practical and publicly accountable, especially in government or highly regulated settings.
- **Agent-analytic + user-centred:** useful when the key behavioural patterns come from ongoing interaction between people and AI systems, rather than from either side alone.
- **Intervention-centred + frame-critical:** useful when AI is being used to make behavioural interventions more targeted or scalable, but the assumptions behind the intervention and the goals of optimisation also need examination.
- **Governance-centred + frame-critical:** useful when AI systems need to be examined not only for oversight and accountability, but also for the assumptions, value choices, and problem definitions shaping the governance response.

The table below develops these pairings in more detail. It shows the kind of question each combination is well placed to ask, what it tends to focus on, what it is likely to contribute, and which further orientation may strengthen it. This is meant as a way of seeing which combinations create useful professional range, rather than as a complete career taxonomy.

The value of these pairings is that they make professional range more concrete: a secondary orientation does not replace your default one - it gives you another way to notice what your usual lens may leave in the background. In practice, that can mean knowing when to bring in another collaborator, when to adjust your own framing, or when to translate your work differently for a particular organisation.

Core question	Focus of attention	Likely contribution	Pair with
User-centred + governance-centred			
How is AI reshaping human judgement, behaviour, and agency, and are our institutions equipped to notice, evaluate, and govern those effects?	Lived behavioural effects, trust, dependence, vulnerability, distribution of burdens, whether governance reflects actual use, public accountability.	Connects human experience to institutional responsibility. Keeps governance grounded in actual use and behavioural consequences.	Deployment-centred or design-centred
Design-centred + deployment-centred			
How can AI systems be designed so they fit human cognition and work within real organisational routines, incentives, and constraints?	Cognitive fit, usability, workflow integration, stakeholder buy-in, capability, operational friction, implementation constraints.	Connects product thinking to real-world adoption. Helps make systems usable, embeddable, and resilient in practice.	Governance-centred or user-centred
Intervention-centred + frame-critical			
How can AI extend the reach, precision, or scale of behavioural interventions, and what assumptions, values, and risks are carried by that promise?	Tailoring, prediction, scaling, optimisation, hidden values, and the assumptions and values behind behavioural optimisation	Keeps applied work ambitious while examining the assumptions behind optimisation. Useful where AI promises to extend behavioural science at scale.	Governance-centred or user-centred

Agent-analytic + user-centred			
What recurring patterns emerge when AI systems and people interact over time? How do those patterns shape system behaviour and human reliance, adaptation, or judgement?	Repeated human–AI interaction, recurring system behaviour, adaptation on both sides, reliance, over-reliance, interaction effects, and patterns that only become visible over time.	Supports the study of human–AI systems as interaction patterns rather than as separate human or machine behaviours.	Governance-centred or deployment-centred
Governance-centred + deployment-centred			
What would it take to translate AI governance principles into workable routines, responsibilities, and oversight practices?	Whether rules and responsibilities match how the organisation actually works, accountability, compliance behaviour, stakeholder incentives, practical oversight.	Turns governance principles into procedures, roles, and organisational realities. Reduces the risk that governance remains abstract.	User-centred or frame-critical
User-centred + design-centred			
How should AI systems be designed in light of how they affect people's judgement, behaviour, trust, and sense of agency?	Lived effects, usability, cognitive fit, friction, trust, interface structure, behavioural responses to design choices	Connects design decisions to actual user conduct. Useful for human-facing services where imagined users are a poor substitute for real ones.	Governance-centred or deployment-centred
Governance-centred + frame-critical			
What kind of arrangement of rules, responsibilities, and power is being built around AI? What political choices are hidden by the way the issue is framed?	Legitimacy, power, accountability, how political choices become treated as technical or managerial questions, framing, whether decisions can be explained and defended publicly.	Useful for catching mismatched problem categories, value choices hidden in technical language, and oversight that looks accountable without changing decisions.	Deployment-centred or user-centred

Potential hybrid roles

Hybrid orientations often become most useful when they can be translated into roles, teams, or organisational problems that other people recognise. The table below sketches how some of these pairings might show up as recognisable forms of work in practice. These are not fixed career paths, and many of the role titles may never appear formally in organisations. They are examples of the kinds of work that can emerge when orientations are combined, from product research and implementation to public-interest governance, model evaluation, and responsible innovation.

You can treat this more as a translation aid than a forecast. Even if a role does not formally exist under these names, the underlying need may still appear under more familiar labels such as product strategy, service design, assurance, research, policy, or transformation. These examples are intentionally approximate and somewhat speculative – many organisations may never use this language directly, and some may end up compressing several orientations into a single vague role title because apparently “AI lead” is expected to contain multitudes.

The intended value of the table is that it helps you recognise what kind of work is being described beneath the label: whether the role is mainly about understanding users, designing systems, embedding tools, evaluating system behaviour, scaling interventions, governing risk, or questioning the assumptions built into the frame itself.

Combination	Possible roles	Potential settings
User-centred + governance-centred	Trust and safety researcher, public-interest product researcher, consumer protection researcher, responsible innovation lead	Public services, regulators, health tech, education tech, civic tech
Design-centred + deployment-centred	AI product lead, service designer, enterprise transformation lead, implementation strategist	Enterprise AI, digital transformation, SaaS, public service reform
Intervention-centred + frame-critical	Behavioural intervention lead, digital interventions strategist, responsible experimentation lead	Health, welfare, education, philanthropy, policy labs
Agent-analytic + user-centred	Human-AI interaction researcher, conversational AI evaluator, model behaviour researcher, sociotechnical safety researcher	AI labs, HCI groups, model evaluation teams, safety research teams
Governance-centred + deployment-centred	Responsible AI operations lead, assurance lead, policy implementation manager	Large firms, regulators, standards and compliance functions
User-centred + design-centred	Behavioural product designer, human-centred AI researcher, UX strategy lead	Product orgs, design teams, AI-enabled services
Governance-centred + frame-critical	AI governance strategist, public-interest AI adviser, responsible AI policy lead, sociotechnical risk analyst	Regulators, think-tanks, responsible AI teams, policy institutes, standards bodies

How the orientations might translate into working styles

The orientations describe different ways of seeing the BeSci-AI landscape. This final section aims to illustrate how those orientations may appear in practice: where someone begins, what they pay attention to first, and what kind of progress they are likely to recognise.

People with similar titles can approach the work in very different ways, especially in a field that is still taking shape. Paying attention to working style can make those differences easier to notice in collaborations, meetings, job descriptions, and project briefs.

N.B. The descriptions below are phrased as tendencies, not claims about everyone working from that orientation.

Orientation	Working style
User-centred	Begins with what people experience and do when they use, rely on, or are affected by AI. Looks at how AI changes judgement, confidence, attention, trust, dependence, and behaviour over time. Typical question: are those effects being taken seriously enough?
Design-centred	Begins with the system, interface, or service as something that can be changed. Looks at whether the design fits how people understand information, make decisions, notice risk, and recover from mistakes. Typical question: what should be redesigned so the system works better for human users?
Deployment-centred	Begins with the organisation around the system. Looks at workflows, routines, incentives, skills, responsibilities, and operational constraints. Typical question: what happens when the system reaches real teams, existing processes, and everyday pressures?
Agent-analytic	Begins with what the AI system does in use, especially over time and across situations. Looks for recurring patterns in outputs, responses, adaptation, and interaction. Typical question: what does the system do across contexts, beyond what it was designed or benchmarked to do?

Intervention-centred	Begins with the behaviour an intervention is trying to change. Looks at targeting, tailoring, feedback, experimentation, delivery, and measurable outcomes. Typical question: could AI help an intervention reach the right people, adapt to context, or produce stronger effects?
Governance-centred	Begins with responsibility around the system. Looks at oversight, accountability, institutional roles, distribution of harms and benefits, and safeguards. Typical question: who is responsible, who is affected, and can the arrangement be explained, defended, and governed?
Frame-critical	Begins with how the problem has been described. Looks at assumptions, inflated claims, mismatched categories, hidden values, and questions that have become harder to ask. Typical question: what does the framing make obvious, what does it make harder to see, and who benefits from describing the problem this way?

Closing thoughts

AI is already changing how people work, decide, delegate, learn, trust, and share responsibility. These changes are often discussed as technical developments, organisational transformations, or governance challenges, but they are also behavioural changes. They depend on what people understand, what they notice, what they are willing to trust, what they are asked to monitor, what they feel able to challenge, and how responsibility is distributed when decisions are partly shaped by systems that people may not fully understand.

That means behavioural science has something important to contribute - not just as an afterthought once the technology is built, and not only as a way to improve adoption or make interfaces more engaging, but as a way to ask better questions about what these systems are being asked to do, what they are changing, and what assumptions about human behaviour are being built into them.

What are people being asked to rely on? What kinds of judgement are being delegated? What forms of oversight are realistic? What assumptions about motivation, attention, agency, trust, or responsibility are being built into products, policies, and organisational processes?

The answer will not come from one kind of behavioural scientist or one version of behavioural expertise - the work ahead is too varied for that.

AI is changing individual judgement and confidence, but it is also changing how work is organised, how services are delivered, how institutions assign responsibility, and how public trust is won or lost. Different behavioural scientists will enter through different parts of that problem. One person may be most useful close to lived experience, noticing how reliance or cognitive effort changes through repeated use. Another may be needed where those effects have to be translated into design, implementation, oversight, or assurance. Others will work further upstream, questioning the assumptions built into the problem before a system, intervention, or governance model is treated as the obvious answer.

This is where understanding your professional orientation becomes useful. When a field is still forming, it can be difficult to know whether your existing skills count, where your experience belongs, or how to explain the contribution you might make. You may recognise that AI raises behavioural questions, but still struggle to see whether your place is in research, product, policy, governance, design, implementation, assurance, strategy, or somewhere that does not yet have a stable name. That uncertainty can make people hesitate, even when their perspective is needed.

The orientations in this guide are meant to make those possible routes easier to see – not boxes to climb into and defend against intruders. They are ways of noticing where your attention naturally goes first, what kinds of problems you are drawn to, and what you may need from people who see the landscape differently. A user-centred orientation may not notice the same

things as a governance-centred one, and a deployment-centred person might not start from the same place as a frame-critical one. That difference is useful, provided people can recognise it before they start mistaking different starting points for disagreement, incompetence, or obstruction.

For individuals, this kind of clarity can help turn a vague sense of relevance into a more concrete professional direction. It can help you say, with more confidence, "this is the kind of AI-related work where my behavioural expertise matters."

For teams and organisations, it can make capability gaps easier to see. A project may have design expertise but no one thinking seriously about implementation. It may have governance principles but no one asking whether those principles can survive contact with real organisational routines. It may have user research but no one connecting lived effects to institutional responsibility. The problem is rarely that one orientation is missing from a diagram. The problem is that a necessary form of thinking may be missing from the work.

The work is already moving faster than the language used to describe it. Waiting until everything is neatly defined before getting involved would be tidy, but it would also mean arriving late to questions that are already shaping workplaces, public services, institutions, and everyday decision-making. The more useful task is to enter with enough clarity to know what kind of contribution you are making, what assumptions come with it, and who you need alongside you.

The purpose of this guide is to give you a way to begin working out where your experience, instincts, and questions might fit. The hope is simple: that people who sense they have something to contribute do not have to spend years working out where the door is.