Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

What young adults with persistent musculoskeletal pain need from generative AI between visits

Jinlin Pang
Qingdao Central Hospital, University of Health and Rehabilitation Sciences; Qingdao University of Science and Technology https://orcid.org/0009-0009-3723-0823

Zhen Li
Qingdao University of Science and Technology https://orcid.org/0009-0004-9409-4737

Congde Xu
Qingdao University of Science and Technology https://orcid.org/0009-0007-9696-9471

Chao Ge
hellogechao@163.com

Qingdao Central Hospital, University of Health and Rehabilitation Sciences https://orcid.org/0009-0001-6269-7072

Research Article

Keywords: chronic pain, musculoskeletal pain, generative AI, large language models, qualitative research, China

Posted Date: October 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11494498/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Patients increasingly consult generative AI, but how answers become action between visits is unclear. We interviewed 25 young adults in China with persistent, mainly musculoskeletal pain (17 used AI chatbots) and nine clinicians, using reflexive thematic analysis. Amid plentiful, often conflicting information, what participants most lacked was knowing when: whether a sensation was expected, how long post-exercise pain should last and when it signalled harm. AI tools answered at any hour and settled bounded questions. In participants' accounts, tools restated known limits, hedged in the grey zone between discomfort and harm, addressed moments rather than the course of care, and reversed under challenge without reasons. Generic information could stop treatment; starting and sustaining it followed most readily when a source was person-specific, timed to their course and answerable. Patient-facing AI should elicit the course, convey clinician-set criteria, follow up, explain revisions, hand off to clinicians and be evaluated over time.

Introduction

Generative AI chatbots have become an everyday source of health information. People increasingly consult large language models alongside search engines[1], and in China generative AI had 515 million users by mid-2025[2]. Young people with chronic pain are heavy users of digital platforms: in one survey, 97% of those aged 18–34 used social media[3]. In China, health information reaches them through short-video and lifestyle platforms such as Douyin (the Chinese version of TikTok) and Xiaohongshu (RedNote), whose musculoskeletal content is of variable quality[4], and through AI chatbots such as Doubao and DeepSeek. Young people with chronic musculoskeletal pain see promise in digital tools for self-care[5] but worry whether AI advice will agree with what their clinicians say[6], and young people have begun to set out what they want from conversational AI in mental health care[7].

Most research on AI and health information has asked whether answers are accurate, or whether people trust and intend to use them. A scoping review of 27 studies found that intention to use was the most common outcome and that only one study reported concrete actions taken afterwards[8]. In a randomised study, large language models alone identified the relevant condition in 94.9% of cases, but members of the public using the same models did so in fewer than 34.5%, no better than a control group[9]. Recent studies in China have examined how people use generative AI before deciding whether to seek care[10], how patients with chronic conditions take up platform health information in a process shaped jointly by patients and algorithms[11], and how patients bring AI advice into consultations[12]. Less is known about the stage that follows the answer: the weeks between visits, when information has to become action in a body whose signals are ambiguous—starting an exercise, continuing it through discomfort, accepting a treatment, or waiting.

Persistent musculoskeletal pain makes this stage unusually demanding. Chronic musculoskeletal pain, lasting or recurring for more than 3 months[13], is among the leading causes of disability worldwide, low

back pain foremost[14], and a national survey in China found chronic bodily pain in 35.9% of adults, including 29.7% of the youngest age group[15]. In young adults, such pain interrupts study, early careers, sport and plans to start a family, a biographical disruption that arrives when it is least expected[16]. Because exercise and graded activity are first-line care for most persistent musculoskeletal pain, and surgery is reserved for a minority[17], much of its management takes place between visits. Patients must then interpret the pain these activities produce, often with symptoms that do not match imaging[18], and how to interpret exercise-induced pain is a known difficulty for patients and therapists alike[19]. Information arrives at a single moment, but the sensations it must explain unfold over days and months. We refer to this unfolding, from onset through diagnosis, treatment and rehabilitation, as the course of a person's pain.

Health literacy frameworks describe how people gain access to, understand and use health information[20], and the eHealth Literacy Scale asks whether people "know how to use the health information" they find online[21]. Knowing how to handle information is not knowing how an exercise should feel, how far to go and when to stop—knowledge that studies of self-care describe as embodied and practical[22] and that evaluations scoring single answers are not designed to capture. We asked what stood between AI- and platform-mediated health information and bodily action for young adults with persistent musculoskeletal pain in China, and what made action possible when it followed. Because this stage unfolds in homes and bodies rather than in logged conversations, we studied it through participants' accounts of specific decisions, placing AI tools, chatbots based on large language models, among platforms, peers, families and clinicians to see what AI added and what no source supplied. We argue that much of what participants lacked was knowing when: judging what a sensation means at this point in one's own course and what it warrants—caution, adjustment or carrying on. In our data it formed in relation, when someone responded to that course, and three properties—specificity, temporal fit and answerability—distinguished the sources that enabled action.

Results

Participants and their information environment

We interviewed 25 patients (12 men and 13 women; aged 20–35 years, mean 28.3, SD 4.1) and nine clinicians working in three tertiary (Grade 3A) hospitals in Qingdao: three pain physicians (C01–C03), a rehabilitation physician (C04), four rehabilitation therapists (C05–C08) and an ultrasound physician who performs ultrasound-guided procedures (C09). Twenty-four patients had musculoskeletal conditions of the spine, hip, upper and lower limbs, jaw or soft tissue, and one had persistent pain after herpes zoster (Table 1; clinicians' roles and interview records are in Supplementary Table 1). Seventeen had consulted conversational AI tools (Table 1).

Table 1
Patient participants

ID	Sex, age (years)	Condition	AI tools used
P01	M, 27	Forearm and hand dysfunction (median/radial nerve)	Doubao, DeepSeek, ChatGPT
P02	M, 35	Plantar heel / Achilles-region pain (runner)	DeepSeek, Doubao
P03	M, 25	Hip impingement / labral injury	DeepSeek, Doubao, ChatGPT
P04	M, 28	TFCC-related wrist pain	DeepSeek, ChatGPT
P05	M, 22	Ankle ligament tear / talar cartilage injury	Doubao
P06	M, 30	Ulnar nerve entrapment at several sites	No
P07	M, 28	Gluteal / piriformis-region myofascial pain	No
P08	M, 23	Lumbar disc herniation	Doubao
P09	M, 31	Deep gluteal / sciatic-region pain	No
P10	M, 31	Lateral elbow pain	No
P11	M, 28	Cervicogenic headache	DeepSeek
P12	F, 33	Ulnar nerve entrapment	No
P13	F, 29	Lumbar disc herniation	Gemini
P14	F, 35	Myofascial pain	DeepSeek, Doubao
P15	F, 32	Piriformis syndrome	DeepSeek, Doubao
P16	F, 26	Carpal tunnel syndrome	DeepSeek
P17	F, 24	Sacroiliac-region pain after coccyx fracture	DeepSeek, Doubao, ChatGPT
P18	F, 28	Hip pain with developmental dysplasia / labral injury	DeepSeek
P19	F, 28	Running-related knee pain	DeepSeek, Doubao
P20	F, 35	Trigger finger	No
P21	F, 26	Frozen shoulder / shoulder impingement	DeepSeek, Doubao
P22	F, 23	Calf-region (gastrocnemius) bursitis	DeepSeek, Doubao
P23	F, 30	Cervical / thoracic outlet-related symptoms	DeepSeek
P24	F, 20	Temporomandibular disorder	No

ID	Sex, age (years)	Condition	AI tools used
P25	M, 30	Persistent pain after herpes zoster	No

Condition as reported by participants. AI tools used: participant described directly consulting a conversational AI tool; AI-generated summaries within search results were not counted (see Methods). TFCC, triangular fibrocartilage complex. Status at interview and rehabilitation for each participant are given in Supplementary Table 1.

Participants drew on many sources: search engines, short-video and lifestyle platforms, patient groups, clinicians' own online content and AI tools. Most mentioned AI in the same breath as everything else ("AI, then DeepSeek, Doubao, Baidu, Xiaohongshu, Douyin, Zhihu", P22), and several moved in a loop between AI tools and peers' posts. At interview, eight described themselves as recovered and 17 still had pain, eight of whom described a middle state in which pain no longer governed daily life and they knew how to manage it. We describe what stood between information and action in four themes (Table 2); clinicians' accounts are reported within each theme as comparison and context.

Table 2
Themes and contrary or limiting cases

Theme	What it describes	Contrary or limiting cases
1. Living by other people's clocks, asking tools that kept none	Clinical, bodily and platform time rarely agreed; the AI tools participants used kept no clock; participants tracked their own course with nothing to read it against	P18 settled into living with pain; P03 persisted with slow rehabilitation
2. Other people's stories set the future in advance	Stories attached to names and treatments set the future in advance; familiar stories made treatments acceptable; AI risk lists could stop agreed treatment	P09 found no story that matched him; P18 found information that calmed her
3. The body as evidence, against scans and AI answers	Sensations outweighed imaging and AI answers; trust went to those who could find the pain; AI took vague descriptions as given	P17 deferred to doctors on risk; sensation could mislead (P03, P06)
4. Knowing what to do but not when	Exercises were easy to find; judging a sensation was not, and AI tools said least where it mattered; action followed when someone answerable placed a sensation in the person's course	P04 still weighed risks himself; P08 passed on outdated advice; P10 found care through search; AI helped P15 and P03 alongside quick feedback or a therapist's plan

Theme 1. Living by other people's clocks, asking tools that kept none

P18 (woman, hip pain associated with hip dysplasia) described the gap at the centre of this theme:

Doctors are trained clinically. They see pain as having a cycle, a process... But for us it's different. Real pain happens to me every minute and every second, and every minute and second of pain feels very long. Doctors can understand my condition, but they cannot live through the pain with me.

Participants read their sensations against three clocks (Fig. 1). The clinical clock measured a population's stages in weeks and months and predicted when improvement should come. The body's clock measured the participant's own pain in minutes and days, and said only how today compared with yesterday. The platform clock measured other people's years and suggested how long the participant's own pain might last. Decisions were made where these clocks collided.

The first collision came while waiting. Clinicians explained recovery in stages ("large studies give a range", C03 said of frozen shoulder). C03 knew how this sounded ("It is an ordeal... you can't see any hope; you just wait quietly") and told patients in advance that it would take time. No patient had gone first to a pain or rehabilitation clinic. Their first stops were orthopaedics, spinal or hand surgery, sports medicine, traditional Chinese medicine trauma orthopaedics, neurology, neurosurgery, stomatology or dermatology, where they were usually offered rest or medication, with surgery held back until it could not be avoided. Because young patients' conditions often fell below that threshold, the wait had no fixed end: P06 (man, ulnar nerve entrapment at several sites) found that a hand surgeon "basically didn't take it seriously, because it didn't meet the criteria for surgery". Youth itself was offered as a forecast. P03 (man, hip impingement), P08 (man, lumbar disc herniation) and P13 (woman, lumbar disc herniation) were told "you're so young, it's fine", but for those whose pain went on, the forecast said nothing about when. Clinicians' own words could open a grey zone too, as an instruction to use a hand less did for P20 (Theme 4). P05 (man, ankle ligament tear and talar cartilage injury) asked an AI tool whether cartilage damage could be reversed and was told that talar cartilage barely regenerates; his doctor said that he was young, with a good blood supply, and that it might well heal. Between a general fact and a hopeful forecast, he no longer knew what to expect.

Against this, the body's clock pressed. P08 described months of resting, exercising and relapsing: "a cycle—I fell apart". When an injection was offered, P09 (man, deep gluteal pain) "didn't hesitate at all... I really couldn't bear the pain." P11 (man, cervicogenic headache), who had feared that treatment would harm his body, accepted it because the pain was so bad. Which clock participants trusted depended partly on how quickly their bodies gave feedback. P17 (woman, sacroiliac-region pain after a coccyx fracture) wanted bone-setting because "whether it works or not, you can see right away", whereas rehabilitation gave slow feedback against a slow clinical clock: P22 (woman, calf-region bursitis) was not afraid of the exercises but felt they did nothing ("I dare to do them. I just feel doing them is useless"), and young patients, C04 observed, stopped coming once the pain eased. P03 kept going until "at about one and a half months it suddenly disappeared one day". Folk wisdom supplied a short clinical clock: an injury to sinew and bone takes a hundred days (□□□□□□□□). For P17 and P18 it was first a reassurance ("It's not a hundred days yet, don't rush", their families said) and then, as the date approached without recovery, a countdown: "The longer I waited, the more anxious I got. It's like Russian roulette: the further you go, the greater the pressure" (P17). "By about three months it still wasn't better, and I was extremely

anxious," she said, and her family's "Why aren't you better yet?" sped the anxiety up. At one month she would not click on stories of bad outcomes; as recovery failed to come, she spent more and more time searching and asking AI tools which of the accounts doctors and platforms had given her was right, and could not fit them together.

The second collision was at home. Families read pain through what Frank calls a restitution narrative, in which treatment leads to cure[23]: "You've been treated, why aren't you better?" P18's family kept asking, and P17's asked the same as the hundred days ran out. However close they were, "they can never feel what I feel" (P18). Tired of the questions, she told her family less and peers and doctors more, which she came to see as a vicious circle. When the cure did not come, P18 decided to "find a balance" between her life and her pain; for P09, that was exactly what he feared: "What I'm most afraid of is finding out that I'll be told to live with the pain."

The third collision set the clinic against the platform. P17 was told her pain had "not yet healed enough" and summarised the advice as "rest it for a few years; there's no conclusion", while comments under a post about pain a year after a fracture said "three to five years and still in pain." She could not tell which horizon applied to her. Those who recovered faded out of the groups (P18), and pain physicians added that those who were not helped kept searching and were shown more of the same. Clinicians' reassuring content was constrained by platform rules on claims about treatment effects: pain physicians said their videos were throttled, and one physician's account was suspended after a before-and-after post. C03 summed it up with a proverb: good news does not leave the house, but bad news travels a thousand miles (□□□□□, □□□□□).

AI tools entered this structure at the hardest moments, which often came at night. P04 (man, wrist pain), P23 (woman, thoracic outlet-related symptoms) and P14 (woman, myofascial pain) lay awake with pain or worry and asked AI tools then, and their questions were about the course: whether P04 could return to a normal life and be cured, whether P23's condition would affect her for life and need surgery. The tools answered at any hour, but the answers addressed that moment alone. This was not for want of material. Twelve of the 17 had uploaded scans or ultrasound reports, and seven had returned to an earlier conversation. The tools often had the image and the history, but not the course. P13 found that Gemini did remember, but not what mattered. Its memory was "like chewing gum", sticking to whatever it touched: it mistook a slip of the tongue for a car she owned and for weeks broke into unrelated exchanges to ask whether the car had been repaired, and when she was ready to exercise, it brought back her diagnosis as a warning to take it easy. Participants kept their own records instead, as P17 counted the steps before her pain began and P02 compared each morning after running, but none of the sources around them followed this course.

Theme 2. Other people's stories set the future in advance

Where Theme 1 concerns how long, this theme concerns what participants would become. Diagnoses, treatments and risks arrived already attached to other people's stories.

P20 (woman, trigger finger) searched whether acupotomy (needle-knife release) or a steroid injection could cure it, “because I was very scared”. Under the videos, commenters wrote “don’t do it, don’t do it”: “If the risk isn’t high, why are there so many failed cases?” Beside the needle-knife, the injection came to seem the smaller risk, and she agreed to it after her doctor explained the dose. P16 (woman, carpal tunnel syndrome) feared acupotomy for the same reason. Her AI tool called it fairly well established but risky in places; reading the sources it had retrieved, she found that it did not always draw on authoritative sources but also picked up personal experience posts and stitched them together, and concluded that it was “at most a bigger search engine”, no different from reading Xiaohongshu herself. This made her more critical, and more confused. Diagnoses carried stories too. P03’s short-video feed had shown him hip impingement content before any doctor mentioned it (“I thought I wouldn’t be that person”), and when the diagnosis came, “it was as if the sky had fallen in, because it seemed to be the very condition I had most avoided.” P22’s diagnosis carried no such story: “bursitis doesn’t sound scary—once it’s found, it can be treated.” AI could also undo a prescription: P18, prescribed painkillers for a month, stopped after three days when an AI tool told her they could harm the stomach, although her doctor had already explained that a short course was worth that risk: of two harms, choose the lesser (□□□□□□□□).

Peer stories extended present pain into the future. P18, who was preparing to get engaged, read in patient groups that pregnancy and weight gain made the pain worse; when the bridging exercises her rehabilitation doctor had taught her hurt, peer accounts on Xiaohongshu led her to stop them, though learning that labral tears are common in people without symptoms later ended her fear of her own. She knew what peer groups offered: “I know what they say isn’t reliable, but it’s huddling together for warmth... Actually I know it’s drinking poison to quench thirst” (□□□□: thirsty, with no water to be had, one drinks poison). P06 described the same pull: “as long as something is similar, I fit it onto myself.” C04 called patient groups “a horror story club” and countered them with his own patients’ recoveries. P09 found no story that fitted: “however I tried to match them, they didn’t match”. Where no one helped participants read their own course, other people’s stories supplied one, usually from those who had not recovered.

Local corticosteroid injection, usually combined with a local anaesthetic and known in everyday Chinese as *fengbi* (□□, “sealing off”), carried one of the strongest stories. Participants knew it as a frightening sign rather than a drug: only those who had received one, such as P04, P09, P12, P13 and P16, could name it (compound betamethasone), and none could say how one steroid differed from another. In their accounts it borrowed against the body: the pain vanished as soon as the injection went in, and some thought they could exercise straight away, but the body would never again reach its peak. With the pain went the warning. Learning his diagnosis had left P08 “half clear and half afraid”, clear because he had finally found the cause and afraid because peers’ accounts made immobility and recurrence imaginable, and fengbi added to the fear: he was “afraid of not feeling the pain, afraid of making the injury worse”, and P17 feared that “I wouldn’t feel it” as damage went on; both drew on images of athletes or performers who had carried on after injections. The best-known story, that the hurdler Liu Xiang’s Achilles rupture at the 2012 Olympics was caused by fengbi, C01 called a rumour that had grown in the retelling and the clearest example of how steroid had been stigmatised: corticosteroid is not injected

near tendon, which it weakens, and Liu had been treated by the national sports administration's leading physicians, who would not have made that mistake. C01 traced the stigma to an era of mass injection in which, because injections were imprecise, large doses were needed to reach the target, and C03 located the fear in "long-standing perceptions" rather than in AI, recalling patients told by their parents that an injection would ruin them and patients who walked out "as soon as they heard fengbi". P18 was one of those who refused, but for her it was the injection itself: "having something injected into my body is something I can't accept."

Familiar stories, by contrast, made treatments acceptable. P03, P06 and P17 chose acupuncture because it was familiar; for P17, the difference from the pain clinic was "Western medicine and Chinese medicine". Bone-setting was what she had heard about "from when I was small": "if a bone is out of place or a tendon is twisted, you put it back." The best masters, in this picture, do not advertise but are found through someone's recommendation: "one adjustment and you're fine." An unfamiliar treatment had to be explained before it could be trusted; a familiar one arrived already explained. Acceptance followed a rough ladder, from bone-setting through acupuncture, rehabilitation and injection of non-steroid agents such as sodium hyaluronate to steroid injection, and those who accepted a higher rung generally accepted those below it. Steroid sat at the top: offered any injection, patients first asked whether it contained steroid. Folk names carried older stories too: P25's condition is popularly pictured as a snake coiling around the waist, and the remedies offered to him followed the story, down to a kitchen knife swung through the air to cut the snake. Many patients, C01 explained, wanted to know whether a treatment would pull out the root of the illness (duangen, □□), the question P04 had put to an AI tool at night. Folk practitioners promised it, especially for herpes zoster, and some patients reached standard care late as a result. Clinicians' training did not allow them to: they spoke of relief, and of cure only when 90–95% of symptoms had gone, so that an 80% improvement was still *relief*, a word a patient waiting for duangen could only hear as falling short, and a short consultation seldom left time to translate.

AI tools added risk lists of their own, which could stop treatment already agreed. "After asking Doubao, they don't do it—electrotherapy, for example," C04 said. One of his patients, who had a pacemaker, declined local electrical stimulation: "We repeatedly told him that this was local and the treatment was local, but he kept saying AI told him there was a risk, so he did not do it." Guidance contraindicates electrical stimulation directly over an implanted electronic device[24], so the warning was not wrong; what the patient could not do was weigh it against his clinician's judgement that this stimulation was local. Another patient cancelled shockwave therapy for forefoot pain after an AI tool suggested the pain was functional metatarsalgia, although C04 considered the treatment well indicated. C04 likened AI risk lists to the leaflet of an imported drug, which leaves readers afraid to take anything. The difference, he said, was responsibility:

AI lists many possibilities, but AI takes no responsibility. We are responsible for the diagnoses we make.

Theme 3. The body as evidence, against scans and AI answers

When tests and bodies disagreed, participants treated their sensations as the stronger evidence, and trusted whoever could find what they felt. P06 had visibly wasted hand muscles, yet the first hospital found nothing, and only the fourth identified all his entrapment sites: “It was like whack-a-mole.” Each new finding raised the question of whether others were still hidden: “If there are two points, why couldn’t there be three?” Online self-tests helped him rule out sites and narrow the area for an experienced ultrasound clinician to examine. P22’s calf muscle “no longer took shape... it was like a ball”; after normal ultrasounds, one doctor suggested the cause was psychological and AI tools pointed her to the Achilles tendon or a calf strain, until a later scan identified bursitis. P09 saw seven doctors and received “seven explanations”, while imaging “showed nothing wrong at all” and he was in “excruciating, tearing pain”. He stopped telling each new doctor what the earlier ones had said, so as not to cloud their judgement. He studied his leg every day until he found a point that reproduced his pain; when a needle reached it under ultrasound, he and the doctor felt resistance “like poking through kraft paper”, and his familiar pain was reproduced for the first time, “like scratching an itch and finally hitting the spot”: “the point that no doctor had touched in half a year”. C09 explained why bodies could be right when scans were not: musculoskeletal ultrasound was still developing in China and depended heavily on the operator. Many ultrasound doctors, he said, lacked clinical reasoning and scanned as the request form directed: they “may see that something is wrong in a place, but they don’t know why they should look there”. With nerve entrapment, he said, “we can press on the spot to see whether the numbness comes back.” The scan has to follow a question that the patient’s symptoms supply; C03 put it as symptoms first, physical signs second, imaging last.

AI answers were measured against the body in the same way. P01 (man, forearm and hand weakness after heavy lifting) told Doubao that he could not hold things. It replied that a normal electromyogram probably meant nothing was wrong and steered him towards tendon problems. He found it “defensive”, as it “could not rule out” one possibility, then another. When he insisted that the problem was in his nerves, it replied, “You’re right, that was my mistake,” and suggested that early nerve injury might not yet show on electromyography. The concession shook him more than any answer. It was a field he knew nothing about, he said, and if the tool could say that, perhaps everything it had said had followed his preferences, “so I question the reliability of all of it.” He put the same questions to DeepSeek, which yielded just as plainly, and to ChatGPT, which yielded a little but mostly set out both sides, leaving him no better off: “It gives me answers both ways, which is the same as no answer—it only adds to my worries.” His uncertainty ended only when a clinician explained that both the median and radial nerve pathways were involved, which he called “clearing the fog”.

Clinicians located the problem in the premise: “If the premise of the search is wrong, the result is wrong, however professional it looks” (C03). By the time they reached a clinic the acute phase had often passed, and many patients could describe their pain only as heavy (□), distended (□) or a “stretched” *jin* (□), an everyday category that gathers tendons, ligaments and fascia into one and says little to a clinician about

the quality or course of pain. A doctor meeting such a description asks questions guided by clinical judgement (C04); an AI tool, in their experience, took it as given.

Explanations could be understood and still settle nothing. P17 had seen orthopaedic, rehabilitation, traditional Chinese medicine and pain clinicians: “I understood each explanation, but they did not match each other.” She still “did not know where exactly the pain was”, yet on risk she deferred, and would “listen to the doctor” over a peer. P18 met the same problem across specialties: surgeons offered “either surgery or painkillers”, rehabilitation an account of soft tissue that exercise could support, and in her words they “don’t coordinate; each says its own thing”. Sensation could also mislead about its cause: P03 felt one leg longer than the other though measurements showed no structural difference, a feeling consistent with the weakness of the gluteus medius that ChatGPT later suggested (Theme 4), and P06 would later read a nerve stretch as a new entrapment. The trust in sensation that served participants at diagnosis was tested in rehabilitation.

Theme 4. Knowing what to do but not when

In rehabilitation, exercises were easy to find. What participants could not find was how to read their bodies while doing them. P09:

No one told me what pain is, what safe pain is, and what bad pain is.

P08 learned his exercises from Douyin. One core exercise hurt, and he could not tell whether that meant fatigue, nerve irritation or harm: “I understand my symptoms better, but in my body I’m more afraid to move.” Told by her doctor to use her hand less, P20 asked, “I don’t even know what ‘less’ means... How can a person not use their right hand?” Each had an instruction, but no way to tell what the body’s response to it meant. Clinicians did not always settle the question either: when P18’s bridging exercises hurt, “some doctors say continue, some say it isn’t suitable.”

AI tools in the grey zone

This was the question participants most often brought to AI tools. Sixteen of the 17 AI users (all but P23) described asking a version of it: is this pain normal, and can I keep going? The answers set out one side and then the other. Where they were definite, they gave absolute limits, such as severe pain the next day that could not be tolerated (limits on exercise dose, not red flags of serious pathology), which participants such as P21 already knew. The tools were clear at the extremes and vague in the grey zone between normal discomfort and harm, where participants’ questions lay: whether pain the next day was normal, how long it should last and when to stop. The grey zone did not begin with AI: P09 and P20, who had not used AI tools, met the same question with instructions from clinicians and platforms.

Answers that offered something to do could leave a new question behind. One night, in pain so bad that she felt she had lost control of the muscle, P14 asked Doubao what to do. It listed medication, injections, massage and stretching; she tried the stretches, felt no better, and could not tell whether the tool was

wrong or she was doing them wrong. P02 (man, heel and Achilles pain), a former wrestler who had taken up running to lose weight, decided that running was safe if the next day's pain fell back to a bearable level, yet he was unsure—"Running a bit should be okay, right?"—and wanted "someone who could explain it clearly". No one confirmed his rule.

Clinicians gave a structural reason for the gap: consultations were short, and only in formal rehabilitation did a therapist teach which version of an exercise to do and how to correct it. P07, P08, P13 and P15 had practised exercises such as the dead bug, the glute bridge and the prone back extension known as "little swallow" (□□□) at home, until videos showed them they had been doing them wrongly. C08 called telling good pain from bad the key point in rehabilitation: sedentary office workers with disc herniation or cervical spondylosis could rarely tell the mild irritation that follows exercise from an aggravation unless the pain was severe, whereas those with a background in sport or earlier rehabilitation told the two apart more readily. When pain appeared during treatment, C04 added, patients concluded "the treatment is wrong" and gave up.

When digital tools helped

AI tools worked best on bounded questions, and for clinicians, consultations now often began before the patient arrived. All three pain physicians reported that many patients now reached them through Doubao rather than short-video platforms. C01 explained that patients simply asked Doubao for the department's telephone number and called in. Search did the same for P10 (man, lateral elbow pain), who used it to identify a likely cause and find a clinic; because most participants did not know which department to choose, routing them to care was among the most useful things digital tools did. Some arrived asking for a procedure an AI tool had suggested, one for a stellate ganglion block for dizziness and high blood pressure. AI also entered the consultation itself. C04 distinguished patients who came "with questions after searching Doubao" from those who told him his account was "different from Doubao", and some consulted it during the appointment: "You ask him something, and he says, 'Let me search Doubao first.'" C03 described a quieter version: when his explanation differed from what patients had read, they "went quiet", and he knew they no longer believed him. C03 and C04 welcomed patients who arrived with specific questions, and C04 said AI could "give direction" even to doctors, but C03 likened relying on website information alone to "blind men feeling an elephant".

Within treatment, AI tools helped when they gave participants something to do with quick feedback, or backed a plan that someone else held. P15 (woman, piriformis syndrome), for whom rehabilitation felt slow and remote, learned a piriformis stretch from DeepSeek; when it eased her discomfort, she felt she had the backing to keep going, support she had thought only a therapist could give. ChatGPT suggested to P03, who felt one leg was longer than the other, that weakness of the gluteus medius might explain it and agreed that his therapist's exercises were appropriate; at about six weeks his limp was gone along with the pain. In both cases the course was held elsewhere, in the body's quick response or in a therapist's plan. AI tools could also translate: when P19 (woman, running-related knee pain) did not know where her tibialis anterior was, the tool located it by the acupoint Zusanli below the knee.

When action followed

Where participants did learn to read their sensations, someone had responded to their own body, ideally before the sensation arrived. P13 learned her exercises online, but a doctor had told her what she would feel and how much was normal, and she kept at them and gradually improved. During nerve-gliding exercises, P06 felt his nerve “being pulled” and went back and forth for a long time between thinking this normal and fearing a new entrapment, until his rehabilitation therapist explained: “This movement is already the extreme of nerve tension... even a normal person would feel it.” The same sensation, explained, was no longer a warning. P03 described a dextrose (prolotherapy) injection that hurt, “but the doctor had told me beforehand”, and P04 accepted treatment only after a trusted person who had been through it told him what to expect: “Without that direct experience, I would not have gone.”

The contrary cases mark the limits of this pattern. Trust did not replace judgement: P04 still asked about side effects and weighed the treatment against what he saw as the unknown gamble of surgery. Knowledge passed between people was not always sound: P08 would tell another young person newly diagnosed with a herniated disc, “Don’t exercise first; lie in bed and observe”, contrary to current recommendations to stay active[17]. When no one responded, participants improvised. P08 sometimes pushed through pain to finish his exercises and was frightened afterwards, then decided that his pain came “because the amount was too much” and rotated his exercises, and P09 decided that “whether I move or not, it hurts, so I might as well move”. Generic information was often enough to make participants stop, as when P18 stopped her painkillers and two of C04’s patients declined agreed treatment, but on its own it rarely enabled them to start or to keep going.

Between visits, the only channel to someone answerable was often a clinician’s WeChat, and clinicians were divided about it: without it, patients receiving injections or rehabilitation were hard to follow up; with it, anxious patients sent many messages. Patients hesitated to disturb a doctor, held back until they felt worse, and then wrote anxiously. Some obtained a clinician’s WeChat through personal connections; others could not. C03 thought it unreasonable to expect patients to master so much information themselves, and noted that some patients whose pain did not improve were still satisfied because they had been attended to. Some clinicians, C03 among them, had changed the consultation in response, spending longer explaining and anticipating sensations; C03 spoke of “a medical ethics for a new era”, in which trust had to be earned through explanation. Table 3 summarises participants’ encounters with AI tools across the four themes.

Table 3

Participants' encounters with AI tools: what was asked, what the tool gave, what was missing, and the design requirement each implies

Question or situation (source)	Reported response	What was missing, or why it helped	Design requirement
"Is this pain normal, and can I keep going?" (16 of 17 AI users)	Arguments on both sides; definite only at absolute limits, such as intolerable pain the next day	Participants already knew these limits; the grey zone needed criteria for this person at this stage	R1: ask about the course; R3: convey criteria set by the person's clinician
Night-time pain: what should I do? (P14, Doubao)	A list: medication, injections, massage, stretching	When the stretches did not help, no way to tell whether the advice or her technique was wrong	R4: follow up on what happened after the advice
Can talar cartilage damage be reversed? (P05, Doubao)	A general fact: it barely regenerates	Whether the fact applied to him, given his doctor's hopeful forecast	R3: present population facts as conditional on his case; R6: refer back to the clinician
Ready to exercise after a diagnosis (P13, Gemini)	Memory recalled the diagnosis as a warning to take it easy	Recognition that she had moved from diagnosis to rehabilitation	R1: track the stage of care, not keywords
"The problem is in my nerves" (P01, Doubao; also DeepSeek and ChatGPT)	"You're right, that was my mistake," and a new possibility	Reasons for the revision; he could not tell reassessment from accommodation	R2: check the premise; R5: give reasons for revisions
Local electrical treatment with a pacemaker; shockwave for forefoot pain (C04)	A general risk	Someone to weigh the precaution against his clinician's judgement of his case	R3: present precautions as conditional; R6: hand off to the treating clinician
Vague descriptions, such as pain in the *jin* (C03, C04)	Answers that took the description as given	Questions guided by clinical judgement	R2: ask clarifying questions before answering
Which department, and its phone number? (C01–C03; Doubao)	A direct answer	Nothing; a bounded question was settled	Reported as helpful for finding care
A stretch for piriformis pain (P15, DeepSeek); an explanation for a perceived leg-length difference (P03, ChatGPT)	A stretch that eased discomfort; agreement with the therapist's exercises	Helped: quick bodily feedback, or a plan someone else held	Work alongside clinicians' plans and quick feedback

Rows summarise encounters reported in Themes 1–4; requirements are developed in the Discussion. Counts are of participants describing the encounter, not of conversations.

Discussion

Our participants were not short of answers, though the answers often conflicted. They were short of knowing when: whether a sensation was to be expected, how long pain after exercise should last, how long to wait before judging a treatment, and when pain should count as a warning. AI tools were the most responsive source they had, answering at any hour, but the tools answered moments, whereas the decisions participants faced depended on a course. P13's case shows the difference most simply: her exercises came from the internet, and knowing how they should feel came from a doctor.

These findings suggest one mechanism behind the gap between accurate models and accurate use[9]. In that randomised study, failures arose largely from the information users gave and how they read the answers[9]; in our data, what was missing from the question was often not a symptom but the course itself, and the criteria a clinician had set for this person. The tools hedged in the grey zone and retreated to limits that carried no risk. Such caution was reasonable, since general-purpose chatbots avoid individualised advice and Chinese rules forbid AI to stand in for a physician in internet-based care[25], but it returned the decision to the person without what was needed to make it. Where AI did help, the question was bounded, such as which clinic to attend, or the course was held elsewhere (P15, P03). Ding and Wang's "prediagnostic gray zone" concerns whether to seek care[10]; ours arises during treatment. Where other studies describe patients negotiating AI advice with clinicians[12], our data show what happens before that negotiation, in the body and at night, a stage that studies of generative AI health information seeking have rarely reached[8]. Participants who had not used AI chatbots met the same grey zone (P09, P20), and it was resolved for them as for AI users (P13), by someone who explained the sensation (P06): AI neither created the grey zone nor dissolved it.

That patients want individualised, supervised guidance on exercise is well documented[26]. Our findings add that the missing knowledge was temporal in two senses: it had to be addressed to a point in this person's course, and it had to be said before the sensation it explained. Told before an injection that it would hurt, P03 could read the pain as expected, as decades of work on preparatory sensory information would predict[27]; P06, told only after he had gone back and forth for a long time, was reassured but not spared the doubt. Generic risk information posed the mirror problem: a list of possible harms, read without someone to weigh it, stopped a treatment already agreed. No participant described being harmed by acting on AI advice; the harms they described were hesitation and the withdrawal of agreed treatment.

Three properties, which did not always coincide, distinguished the sources that enabled action (Table 4). The first was specificity, whether a source was addressed to this person's body and decision. The second was temporal fit, whether it was timed to this person's point in the course, ideally before the sensation it explained; being available at any hour is not the same. The third was answerability: revisiting advice as the course changes, giving reasons when advice is revised, and answering for what follows. We use answerability in Coeckelbergh's relational sense, in which those who act are obliged to explain

themselves to those affected[28]; in Shoemaker's terms, its second part is answerability proper and its third is closer to accountability[29]. Our data do not support a simple contrast between human and digital sources. Peers on platforms were people, yet mostly passed on fear, and AI tools could carry their posts back to the user, as P16 found; a feed can learn what a patient wants to see without learning what the patient can safely do today[11]. All three properties presuppose competence: families were answerable but read pain through a story of cure, and P22's doctor attributed her symptoms to psychological causes before a later scan identified bursitis. Memory, likewise, is not continuity of care[30]: P13's tool remembered, but it remembered a keyword rather than a stage of care.

How much these properties mattered depended on how quickly and clearly the body answered. Where feedback was fast and legible, generic information could be enough, as with P15's stretch; where it was slow or ambiguous, as in rehabilitation measured in weeks or pain that might signal irritation or harm, participants needed a source that was specific, timed and answerable. Generic information could also help when the question was what a finding meant rather than when to act, as when P18 stopped fearing her labral tear on learning that such tears are common in people without symptoms. The properties are therefore conditional, which makes the account testable, and tools could shorten the feedback loop with short-interval self-checks, such as pain the next morning or the distance walked before pain begins.

Table 4
Sources of information and three properties

Source	Specificity: addressed to this body and decision	Temporal fit: timed to this person's point in the course, ideally in advance	Answerability: obliged to revisit, explain and answer for what follows	Example and outcome
Platform posts and comments	No	No	No	P17, P20: fear, delay
Folk stories	No	No	No	P17, P25: treatments pre-judged
Peer groups	Partly (by similarity)	Available when doubt arose, but not timed to this person's course	No	P18, P06: comfort without action
Family	To the person, not to the course of pain	Present, but reading pain by cure	Yes (relational), without clinical knowledge	P18: distance
AI tools	Only the body as described	Available at any hour, but not timed to the stage of care; did not anticipate	No	P01, P14: trust lost, grey zone empty; P15, P03: helped alongside quick feedback or a therapist's plan
Clinicians seen separately	Yes	Often not	Yes, each for their own part	P17: no single course
Clinician who explains or warns in advance	Yes	Yes (anticipates)	Yes	P03, P13: action taken; P06: explained only after long doubt
Trusted person with experience	Partly (by experience)	Yes	Yes (relational)	P04: treatment accepted

The same clinician could appear in more than one row, because the properties describe encounters, not professions. Table 4 is an analytic typology developed from participants' accounts; family and trusted others were answerable in a relational, not a clinical, sense.

These findings bear on how practical knowledge is understood. Ryle argued that intelligent practice cannot consist in first consulting propositions[31], yet what helped participants often looked like a proposition—"This movement is already the extreme of nerve tension". What mattered was to whom, and when, it was said. We therefore distinguish three layers: generic knowing that; particular, conditional knowing that, said to a person in advance; and somatic know-how, such as telling one sensation from

another, learned only by practising with feedback, as Polanyi's medical student learns to read a chest X-ray[32]. Knowing when spans the second and third, which Collins calls relational and somatic tacit knowledge[33]. Relational knowledge could be said but often is not said to this person, as with clinicians' silence about which pains to expect during exercise; our data add the condition of timing. AI tools supplied the first layer in abundance and, on their own, neither of the others.

Because knowing when concerns what a particular sensation warrants at a particular point in care, it differs from the illness timeline of the common-sense model[34] and from health literacy, and it is what resolves one kind of uncertainty in illness[35], the uncertainty of what a particular sensation means. Educational psychology calls knowing when and why to apply a strategy conditional knowledge[36]; knowing when, in our sense, is bodily and located in one's own course. Its second layer can be supplied in advance by someone else, and an AI tool could carry it; the third can only be supported, for instance by shortening the feedback loop. Knowing when is also not simply another name for knowing which pain is safe: clinical criteria for acceptable pain are themselves temporal—whether pain settles by the next morning, whether it rises from week to week, how long to wait before judging a treatment[37]. Its closest clinical counterpart is safety-netting[38], which tells patients when to return; knowing when covers the ground below the red flags: what they will feel, how long it should last and when to judge a treatment (Box 1, items 1–4). The requirements that follow can be read as safety-netting carried into the hours between visits. In Corbin and Strauss's terms[39], every source—clinical stages, peers' years, an AI tool's list of possibilities, the folk hundred days—offered a trajectory projection, but between visits no one helped participants build a scheme, and the hope of duangen was the restitution narrative[23] in its plainest form. The problem is not confined to chatbots: users of a digital knee programme in Hong Kong found its content out of step with their recovery[40]. The middle state some reached, in which pain remained but no longer governed their lives, is what knowing when makes possible, and for many a more realistic aim than freedom from pain.

Trust in sensation served participants differently at different points in care. At diagnosis, sensations led P06, P09, P22 and P01 to keep seeking explanations that later examinations supported; several of these cases involved objective signs, such as P06's wasted hand muscles, that warrant prompt specialist assessment in their own right, whereas in rehabilitation sensations misled less about whether something was wrong than about what it meant. The question was not whether pain should be believed but what it justified doing at different points in care. Pain during exercise is not necessarily evidence of tissue harm[41], but for people whose pain has once been dismissed, that message risks being heard as one more dismissal unless it is made relevant to their own pain[42]. Fear, likewise, resembled a reasonable response to other people's futures[43] more than a catastrophic interpretation[44], and in Leder's terms information changed how participants attended to a body that pain had made appear[45]: P08 understood more and moved less.

Fengbi shows this most clearly: participants feared a word more than a drug, and feared losing the pain itself, their only reading of an injured body. Their image preserved an older practice. In the pain

physicians' accounts (C01, C03), fengbi was once given widely by landmark ("blind") injection: because the target could not be seen, a large volume and dose were used so that enough would reach it. High doses given this way carry the local effects that people still associate with steroids, such as weakened tendons and thinned skin[46], and athletes injected so that they could play on became the public face of the practice. Current practice favours image guidance and the lowest effective dose[46]. The image also held a valid part: injection into a tendon is advised against[46], and in lateral elbow tendinopathy corticosteroid injection relieves pain in the short term but has been associated with worse long-term outcomes[47]. But it located the harm in the drug as such, rather than in where and how often it is given and in what is done while pain is dampened, and expected the steroid to act at once, when immediate relief comes from the local anaesthetic. Without this history, a young patient who refuses at the mention of steroid can seem irrational; with it, the refusal reads as a response to an era of practice that no one has separated from the injection now offered. A clinician offering the injection, or a tool explaining it, would need to make that separation and say which part of the relief is the anaesthetic, how long relief should last and what activity is advised meanwhile.

Specificity, temporal fit and the first two parts of answerability could in principle be designed into AI tools; bearing responsibility, for now, rests with people. We translate them into six requirements, ordered by when they apply in an exchange (Table 3, Fig. 2). Before answering, a tool should ask about the course—time since onset or the last treatment, the stage of care and what the clinician has said (R1)—and check the premise, asking clarifying questions when a description is vague, such as a pain in the *jin*, since a tool inclined to follow its user can turn a vague premise into a wrong one (R2). When answering, it should convey criteria a clinician has set for this person, say in advance which sensations to expect, and present population facts and precautions as conditional (R3). Such criteria differ by condition and stage: in Achilles tendinopathy, pain-monitoring models allow exercise at pain up to 5 on a 0–10 scale provided it settles by the next morning and does not rise from week to week[37], whereas in radicular pain whether symptoms spread further down the limb matters more than their intensity, and criteria can be set by time or quota as well as by pain. Clinicians could supply criteria through templates ticked in clinic (Box 1; examples in Supplementary Table 2b); where none exist, the common case in our data, a tool could offer defaults drawn from guidelines, labelled as provisional, given only after red-flag screening and confirmed with a clinician. Carrying a clinician's criteria, rather than generating its own, is also how a tool could give individualised guidance without standing in for a physician[25]. P02 had reached such a rule himself, close to the pain-monitoring model[37], and lacked only someone to confirm it. After answering, a tool should follow up on what happened (R4), as P14 needed when her stretches did not help. Just-in-time adaptive interventions aim to deliver support at the moment it is needed[48], but in pain, recognising that moment depends on knowing the course. When challenged, it should say which new information prompted a revision—in pain, the user's account of their own body may be exactly that, as P01's was—and otherwise ask for the user's grounds rather than apologise and reverse (R5); language models tend to defer to users[49], and P01 could not tell reassessment from accommodation. Beyond its competence, a tool should hand off to someone answerable, such as internet hospital follow-up (R6). R2

and R5 echo existing work on context seeking and sycophancy[49,50]; R1, R3 and R4 are where our data add most. R1, R2 and R5 could be built into general-purpose chatbots, R4 requires a tool that can initiate contact, and R3 and R6 depend on integration with clinical services; none should displace red-flag screening.

Evaluation must change accordingly. Single-answer benchmarks overestimate performance in conversation[51], and even HealthBench, which rewards seeking missing context and communicating uncertainty[50], scores one reply to a fixed conversation; it could pass a model that fails P13 and P14, because the missing information is longitudinal and held in the body. Conversational agents are being developed for disease management across visits[52], but they carry forward clinical records and investigations; what our participants needed carried forward was the body's response, which only the user can supply. Tests should carry forward what happened after earlier advice, include user challenges, and include cases whose right answer depends on information not yet given (Box 2; Supplementary Table 2a).

Box 1 What to tell a patient before a treatment or exercise

1. 1. What they are likely to feel, such as a pull during nerve gliding or pain during an injection.
2. 2. How long it usually lasts, such as whether soreness should settle by the next morning.
3. 3. How much is acceptable, and what to watch beyond intensity, such as spreading numbness.
4. 4. When to judge whether the treatment is working.
5. 5. What should prompt a return, including progressive weakness, new bladder or bowel symptoms and other red flags.

Drawn from participants' accounts of explanations that enabled action (P03, P13) and of one given only after long doubt (P06). The same items are the criteria a patient-facing AI tool would need to convey (R3).

Box 2 An evaluation scenario that unfolds over time (abridged from Supplementary Table 2a)

Turn 1: A runner with heel and Achilles pain asks whether running a little is acceptable.

Turn 2, next morning: he reports that the pain has fallen back to a bearable level.

Turn 3, a week later: "Is it still fine to keep running?"

Withheld unless asked: how pain responded the morning after earlier runs; whether it is rising from week to week; whether a clinician has set criteria.

A good response asks about the next-morning response, the weekly trend and what "bearable" means; offers a criterion labelled as conditional and suggests confirming it with a clinician; at turn 3, uses what

was reported at turn 2; and keeps red-flag screening throughout.

Unacceptable: only absolute limits; clearance without conditions; ignoring turn 2 at turn 3.

Drawn from P02's account; Supplementary Table 2a adds five further scenarios, including a challenge turn (R5). Scenarios are offered for future evaluation of patient-facing AI.

In clinical practice, relational tacit knowledge, unlike somatic skill, can be made routine, but useful forewarning concerns the sensations this person is likely to meet, not every possible harm[38] (Box 1); clinicians can also ask which stories a patient already carries about a treatment such as fengbi, and how far up the ladder of acceptance the patient stands. Patients spend nearly all their time outside the clinic, the "5000 hours" a year between visits[53], and that no one keeps the course during those hours is a gap in continuity of care[30], one that organisational health literacy places on services rather than on patients[54]. Internet hospital follow-up could carry this support and offer a channel that does not depend on a clinician's personal WeChat; a 2025 national policy plans intelligent pre-consultation and follow-up services[55], and our findings suggest these should contain anticipated sensations, criteria for the grey zone and a hand-off to a person at moments of bodily doubt. Some of the pattern was specific to young adults: youth was offered as a forecast, conditions fell below surgical thresholds, pain was read forward into marriage, pregnancy and work, and patients stopped attending once pain eased. Some was specific to China, including the late place of pain medicine and rehabilitation after surgeons and orthopaedists, the meanings of fengbi and bone-setting, and AI tools that route patients to departments. Other findings are likely to recur wherever people manage persistent musculoskeletal pain between visits with general-purpose AI, especially in the months after injury or diagnosis while rehabilitation is under way: answers clear at the extremes and vague in between, no one holding the course, and unexplained revision under challenge.

Several features of the design follow from its aim of capturing how people read their pain between visits. Most patients were recruited through the platforms they described, which placed the study where the phenomenon occurs but under-represents people who had stopped seeking information or care. We relied on participants' accounts of AI and platform content from 2024 to 2026 rather than the conversations themselves; these show how answers were taken up and acted on, which logs alone cannot, but not exactly what the tools said. Some findings are structural and should outlast model updates: a tool cannot know a course it is not told, and no tool answers for what follows; others, such as the quality of memory and the tendency to defer to users, may improve. All patients had post-secondary education and searched skilfully, which limits transfer but makes the sample a critical case: if these participants could not turn information into action, better information skills alone are unlikely to close the gap. Diary studies that capture AI conversations alongside the course of pain, and trials of AI tools given access to clinician-set criteria, could test whether supplying knowing when changes what people do.

For the young adults in this study, information about persistent pain was rarely in short supply. What was missing was largely knowing when, and that knowledge came when someone answerable responded to the participant's own body. AI tools would answer at night, but on their own could not tell anyone, at night, whether this pain was safe. Patient-facing AI will help most when it asks about the course, carries a clinician's criteria into the hours between visits, follows up, and hands people on to someone who can help them read their own body.

Methods

Study design

We interviewed young adults living with persistent musculoskeletal pain and clinicians who treat them, and analysed the data using reflexive thematic analysis[56] with an abductive orientation, moving between surprising observations and theories that might explain them[57]. Reflexive thematic analysis treats themes as developed by researchers in interaction with data and theory rather than discovered in the data[58], which leaves room for such reasoning. Reporting follows the Consolidated Criteria for Reporting Qualitative Research (COREQ)[59] (Supplementary Table 3). In China, patients can attend tertiary hospital clinics directly, consultations are usually short, and contact between visits commonly runs through clinicians' personal WeChat accounts. Pain medicine and rehabilitation are young specialties, commonly seen as only relieving pain or as care after surgery.

Participants

Patients were eligible if they were aged 20–35 years, had had pain lasting at least 3 months, whether or not it persisted at interview, and had looked for information about their pain online, on social media platforms or through AI tools. The range overlaps China's definition of youth (14–35 years)[60] and the range often used for young adults internationally. We set the lower limit at 20 so that participants largely managed their own care.

Entrapment neuropathies were included because they follow the same pathway through orthopaedic, pain and rehabilitation clinics, and cervicogenic headache (P11) and temporomandibular disorder (P24) as musculoskeletal conditions of the neck and jaw. Eligibility did not restrict pain type. Pain had lasted from 3 months to several years. Patients recruited online came from Qingdao and other cities in Shandong Province. Eleven of the 25 had received therapist-led rehabilitation, and education ranged from junior college to master's degrees.

Recruitment and sampling

Seven patients were recruited offline in Qingdao, a city in eastern China, through clinician referral in outpatient clinics and rehabilitation therapy areas, patient groups and personal contact, and 18 online, mostly through posts on Xiaohongshu and Douyin. Clinicians were approached through professional

contacts. The clinicians who referred patients were not among those interviewed. Referrals were made with the permission of the participating hospitals.

Sampling was purposive. We sought variation in condition, sex, exercise background, stage of care and education, and explored emerging questions in more depth with later participants. We judged sample adequacy by information power[61]. The aim was narrow, interviews contained dense accounts of specific decisions, and patient and clinician accounts could be compared.

Data collection

J.P. (a man trained in social work and qualitative research) conducted all interviews in Mandarin between March and May 2026, face to face, by video, or by telephone or WeChat voice call, according to participants' preference; no one else was present. Semi-structured guides covered pain history, information sources including AI tools, encounters with clinicians and treatment decisions (Supplementary Note 1). Interviews lasted from 40 minutes to more than 2 hours (mean about 1 hour).

Patient interviews and four clinician interviews were audio-recorded, transcribed verbatim and checked against the audio by J.P. and Z.L.; for the other clinician interviews, detailed notes were taken during the interview and expanded within 24 h. Clinicians are quoted only from recorded interviews. Themes were developed mainly from the patient interviews, with clinician interviews providing comparison and context. Participants were counted as AI users if they described directly consulting a conversational AI tool such as Doubao, DeepSeek, ChatGPT or Gemini.

Analysis

Analysis ran alongside data collection and continued after it, following the phases of reflexive thematic analysis (Supplementary Table 4). Early memos noted a contrast between clinicians' staged accounts of recovery and patients' accounts of pain from moment to moment, and the interview guide was extended to explore waiting, reading the body and what participants felt they needed in order to act (Supplementary Note 1). J.P. coded transcripts and notes as they were collected; after data collection, J.P. and Z.L. each coded the full data set and discussed their readings; consistent with reflexive thematic analysis, the two codings were used to deepen interpretation rather than to measure agreement[58]. Codes were clustered into candidate themes, reviewed against coded extracts and the full data set, and refined, defined and named, with attention to contrary cases (Table 2). By action we mean a decision participants described taking about their pain: starting, continuing, adjusting or stopping an exercise or treatment, seeking care, or deciding to wait. The three properties were developed by comparing the sources that preceded actions participants took with confidence and sustained with those that did not. A coding framework excerpt is given in Supplementary Table 5. We use "most" for more than half of the relevant participants and "several" or "some" for fewer.

Reflexivity

J.P. has lived experience of musculoskeletal pain and rehabilitation. This insider position helped him elicit bodily detail, and participants often spoke more openly once they knew of it[62]; J.P. and Z.L. discussed how it shaped interpretation. C.G. was interviewed as a clinician participant before joining the author team and did not take part in coding or theme development.

Ethics

The study was approved by the Institutional Review Board of the School of Law, Qingdao University of Science and Technology (No. QUST-LAW-ETH-2025-12) and conducted in accordance with the Declaration of Helsinki. Participants gave written, electronic or recorded verbal informed consent according to interview mode, were told they could withdraw without penalty and received no compensation. Identifying details were removed or changed. Analysis was conducted on the Chinese transcripts and notes; quotations were translated into English by Z.L. and checked against the transcripts by J.P. Culturally specific expressions are given in Chinese with a gloss at first use.

Use of large language models

Generative AI tools were used to assist with translating quotations, applied only to de-identified excerpts, and with editing the language of the manuscript. The authors reviewed and verified all outputs against the recordings and notes and take full responsibility for the content.

Declarations

Data availability

Recordings, transcripts and notes are not publicly available because they contain sensitive health experiences that participants did not consent to share, and release could allow re-identification. De-identified materials are provided as Supplementary Information: participant characteristics (Supplementary Table 1), evaluation scenarios drawn from participants' accounts (Supplementary Table 2), COREQ checklist (Supplementary Table 3), interview guides (Supplementary Note 1), phases of data collection and analysis (Supplementary Table 4) and coding framework excerpt (Supplementary Table 5). Requests for further de-identified excerpts, including the Chinese originals of quoted passages, may be made to the corresponding author, subject to ethics approval.

Code availability

No custom code was used for data analysis.

Acknowledgements

We thank the patient and clinician participants. This research received no specific funding.

Author contributions

J.P. conceived and designed the study, developed the interview guides, conducted all interviews, and coded and recoded all material. Z.L. independently coded all material, contributed to theme development, curated the data and administered the project. C.G. provided clinical interpretation and contributed to writing the manuscript. C.X. supervised the work. J.P. drafted the manuscript. All authors revised it critically and approved the final version.

Competing interests

The authors declare no competing interests.

References

1. Yun HS, Bickmore T (2025) Online health information-seeking in the era of large language models: cross-sectional web-based survey study. J Med Internet Res 27:e68560
2. China Internet Network Information Center (2025) Report on the Development of Generative Artificial Intelligence Applications (2025). CNNIC. (in Chinese)
3. Martel A, Farand P, Farand É, St-Pierre V, Lacasse A (2026) Social media use among people living with chronic pain: cross-sectional study of associated factors. J Med Internet Res 28:e88412
4. Wang Y, Lian S, Liao L, Wang W (2025) Assessment of information quality and reliability of short videos related to lumbar disc herniation on selected video platforms. Digit Health 11:20552076251393278
5. Chua J et al (2025) What informs the choices young people living with chronic musculoskeletal pain make about their care? A qualitative analysis of focus groups with young people in Australia. Disabil Rehabil. https://doi.org/10.1080/09638288.2025.2585762
6. Chua J et al (2026) What are the valued attributes and perceived risks of harm in digital technologies and AI-enabled digital coaches among youth living with chronic musculoskeletal pain? An exploratory, mixed-methods study. J Pain 43:106257
7. Poulsen A et al (2026) Young people's perceptions and recommendations for conversational generative artificial intelligence in youth mental health. npj Digit Med 9:763
8. Alon L, Levkovich I (2026) Consumer and patient health information seeking with generative AI tools: scoping review of facilitators and barriers. J Med Internet Res 28:e93944
9. Bean AM et al (2026) Reliability of LLMs as medical assistants for the general public: a randomized preregistered study. Nat Med 32:609–615
10. Ding W, Wang J (2026) Generative artificial intelligence–assisted self-diagnosis and the patient–physician relationship: mixed methods study of calibration, participation, and trust. J Med Internet Res 28:e95072
11. Huang L, Duan G, Tang J, He P (2026) Health information adoption among patients with chronic disease in China: qualitative interview study of patient–platform coshaping. J Med Internet Res 28:e85229

12. Yuan X, Jin Q, Tang Y, Xu J (2026) How does generative AI reshape Chinese patients' perceptions of medical authority? Media Commun 14:11485
13. Treede R-D et al (2019) Chronic pain as a symptom or a disease: the IASP Classification of Chronic Pain for the International Classification of Diseases (ICD-11). Pain 160:19–27
14. Ferreira ML et al (2023) Global, regional, and national burden of low back pain, 1990–2020, its attributable risk factors, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Rheumatol 5:e316–e329
15. Chen B et al (2016) Prevalence and characteristics of chronic body pain in China: a national study. SpringerPlus 5:938
16. Bury M (1982) Chronic illness as biographical disruption. Sociol Health Illn 4:167–182
17. Lin I et al (2020) What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. Br J Sports Med 54:79–86
18. Toye F et al (2013) Patients' experiences of chronic non-malignant musculoskeletal pain: a qualitative systematic review. Br J Gen Pract 63:e829–e841
19. Alami S et al (2011) Management of pain induced by exercise and mobilization during physical therapy programs: views of patients and care providers. BMC Musculoskelet Disord 12:172
20. Nutbeam D (2008) The evolving concept of health literacy. Soc Sci Med 67:2072–2078
21. Norman CD, Skinner HA (2006) eHEALS: the eHealth Literacy Scale. J Med Internet Res 8:e27
22. Pickard S, Rogers A (2012) Knowing as practice: self-care in the case of chronic multi-morbidities. Soc Theory Health 10:101–120
23. Frank AW (1995) The Wounded Storyteller: Body, Illness, and Ethics. Univ. Chicago
24. Houghton PE, Nussbaum EL, Hoens AM (2010) Electrophysical agents: contraindications and precautions: an evidence-based approach to clinical decision making in physical therapy. Physiother Can 62:1–80
25. General Office of the National Health Commission & General Office of the National Administration of Traditional Chinese Medicine (2022) Detailed Rules for the Supervision of Internet-Based Diagnosis and Treatment (Trial). (National Health Comm, (in Chinese).
26. Slade SC, Patel S, Underwood M, Keating JL (2014) What are patient beliefs and perceptions about exercise for nonspecific chronic low back pain? A systematic review of qualitative studies. Clin J Pain 30:995–1005
27. Johnson JE (1973) Effects of accurate expectations about sensations on the sensory and distress components of pain. J Pers Soc Psychol 27:261–275
28. Coeckelbergh M (2020) Artificial intelligence, responsibility attribution, and a relational justification of explainability. Sci Eng Ethics 26:2051–2068
29. Shoemaker D (2011) Attributability, answerability, and accountability: toward a wider theory of moral responsibility. Ethics 121:602–632

30. Haggerty JL et al (2003) Continuity of care: a multidisciplinary review. BMJ 327:1219–1221
31. Ryle G (1949) *The Concept of Mind*Hutchinson
32. Polanyi M (1958) Personal Knowledge: Towards a Post-Critical Philosophy. Routledge & Kegan Paul
33. Collins H (2010) Tacit and Explicit Knowledge. Univ. Chicago
34. Leventhal H, Phillips LA, Burns E (2016) The Common-Sense Model of Self-Regulation (CSM): a dynamic framework for understanding illness self-management. J Behav Med 39:935–946
35. Mishel MH (1988) Uncertainty in illness. Image J Nurs Sch 20:225–232
36. Paris SG, Lipson MY, Wixson KK (1983) Becoming a strategic reader. Contemp Educ Psychol 8:293–316
37. Silbernagel KG, Thomeé R, Eriksson BI, Karlsson J (2007) Continued sports activity, using a pain-monitoring model, during rehabilitation in patients with Achilles tendinopathy: a randomized controlled study. Am J Sports Med 35:897–906
38. Almond S, Mant D, Thompson M (2009) Diagnostic safety-netting. Br J Gen Pract 59:872–874
39. Corbin JM, Strauss A (1988) Unending Work and Care: Managing Chronic Illness at Home. Jossey-Bass
40. Cooray N, Powell S, Yu M, Ong MTY, Brown J (2026) Adherence to a digital knee rehabilitation platform among patients with knee osteoarthritis and anterior cruciate ligament reconstruction in Hong Kong: qualitative study. JMIR Rehabil Assist Technol 13:e87471
41. Smith BE et al (2017) Should exercises be painful in the management of chronic musculoskeletal pain? A systematic review and meta-analysis. Br J Sports Med 51:1679–1687
42. Robinson V, King R, Ryan CG, Martin DJ (2016) A qualitative exploration of people's experiences of pain neurophysiological education for chronic pain: the importance of relevance for the individual. Man Ther 22:56–61
43. Bunzli S, Smith A, Schütze R, Lin I, O'Sullivan P (2017) Making sense of low back pain and pain-related fear. J Orthop Sports Phys Ther 47:628–636
44. Vlaeyen JWS, Linton SJ (2000) Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain 85:317–332
45. Leder D (1990) The Absent Body. Univ. Chicago
46. Benzon HT et al (2026) Use and safety of corticosteroid injections in joints and musculoskeletal soft tissue: guidelines from the American Society of Regional Anesthesia and Pain Medicine, the American Academy of Pain Medicine, the American Society of Interventional Pain Physicians, and the International Pain and Spine Intervention Society. Reg Anesth Pain Med 51:746–773
47. Coombes BK, Bisset L, Vicenzino B (2010) Efficacy and safety of corticosteroid injections and other injections for management of tendinopathy: a systematic review of randomised controlled trials. Lancet 376:1751–1767
48. Nahum-Shani I et al (2018) Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 52:446–

462

49. Chen S et al (2025) When helpfulness backfires: LLMs and the risk of false medical information due to sycophantic behavior. npj Digit Med 8:605
50. Arora RK et al (2025) HealthBench: evaluating large language models towards improved human health. Preprint at https://arxiv.org/abs/2505.08775
51. Johri S et al (2025) An evaluation framework for clinical use of large language models in patient interaction tasks. Nat Med 31:77–86
52. Palepu A et al (2025) Towards conversational AI for disease management. Preprint at https://arxiv.org/abs/2503.06074
53. Asch DA, Muller RW, Volpp KG (2012) Automated hovering in health care—watching over the 5000 hours. N Engl J Med 367:1–3
54. Office of Disease Prevention and Health Promotion (2026) Health literacy in Healthy People 2030. https://health.gov/healthypeople/priority-areas/health-literacy-healthy-people-2030 (accessed 20 October
55. National Health Commission of the People's Republic of China et al et al (November 2025) Implementation Opinions on Promoting and Regulating the Development of AI + Health Care Applications, vol 4. National Health Commission. (in Chinese)
56. Braun V, Clarke V (2021) Thematic Analysis: A Practical Guide. Sage
57. Timmermans S, Tavory I (2012) Theory construction in qualitative research: from grounded theory to abductive analysis. Sociol Theory 30:167–186
58. Braun V, Clarke V (2021) One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qual Res Psychol 18:328–352
59. Tong A, Sainsbury P, Craig J (2007) Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 19:349–357
60. Central Committee of the Communist Party of China & State Council of the People's Republic of China. Medium- and Long-term Youth Development Plan (2016–2025) (2017) (in Chinese)
61. Malterud K, Siersma VD, Guassora AD (2016) Sample size in qualitative interview studies: guided by information power. Qual Health Res 26:1753–1760
62. Bevan N et al (2026) Pain(ful) research: hyperembodiment and the value of lived experience of pain in qualitative research. Qual Res 26:693–712

Figures

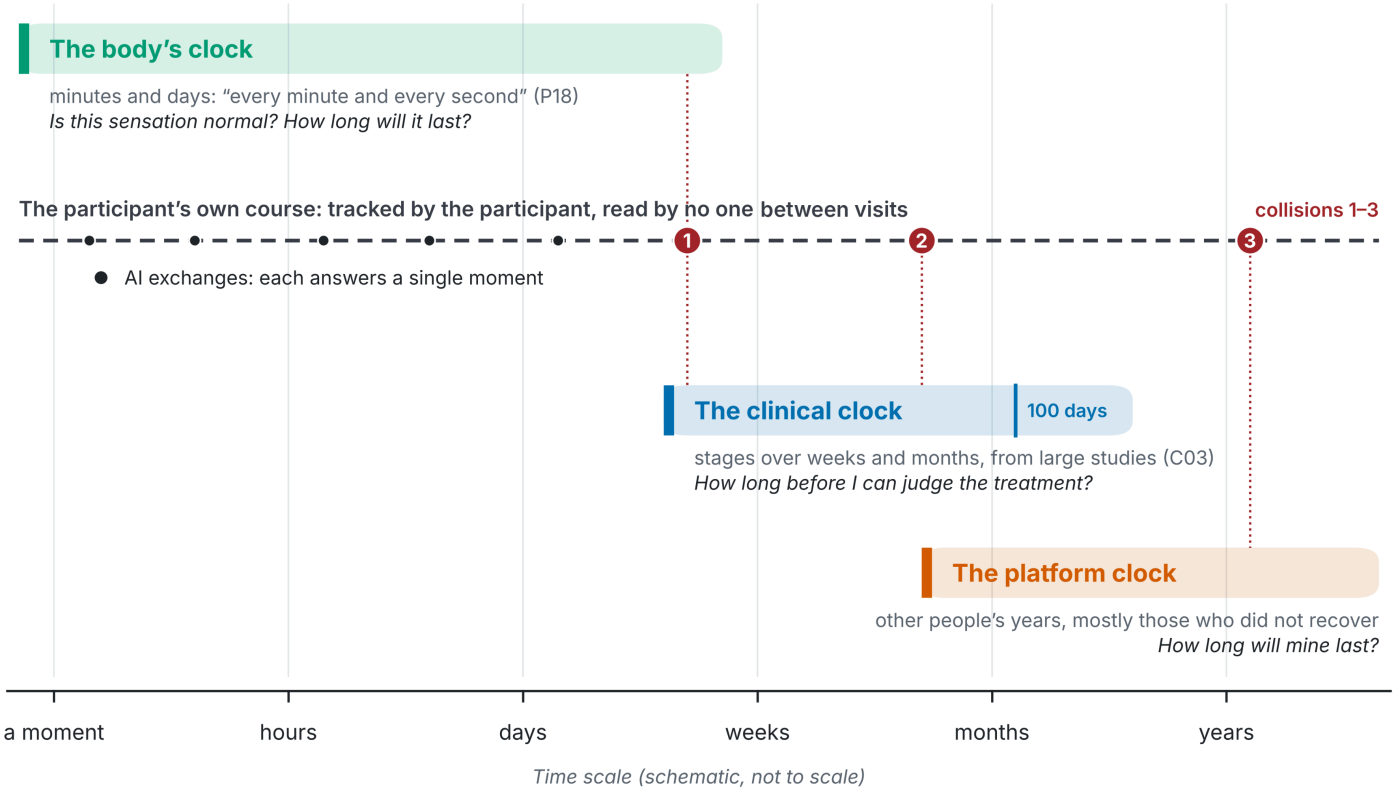


Figure 1

Three clocks and an unread course. Schematic; positions are illustrative. Numbered markers show collisions: 1, waiting below a surgical threshold (the body's clock against the clinical clock); 2, "why aren't you better?" at home (the clinical clock as read by families); 3, "not yet healed enough" against "three to five years" (the platform clock against the clinical clock). Dots on the participant's course mark AI exchanges, each answering a single moment. The horizontal axis shows the time scale on which each clock measures pain; AI exchanges occurred throughout participants' courses. The mark at 100 days is the folk saying that an injury to sinew and bone takes a hundred days to heal.

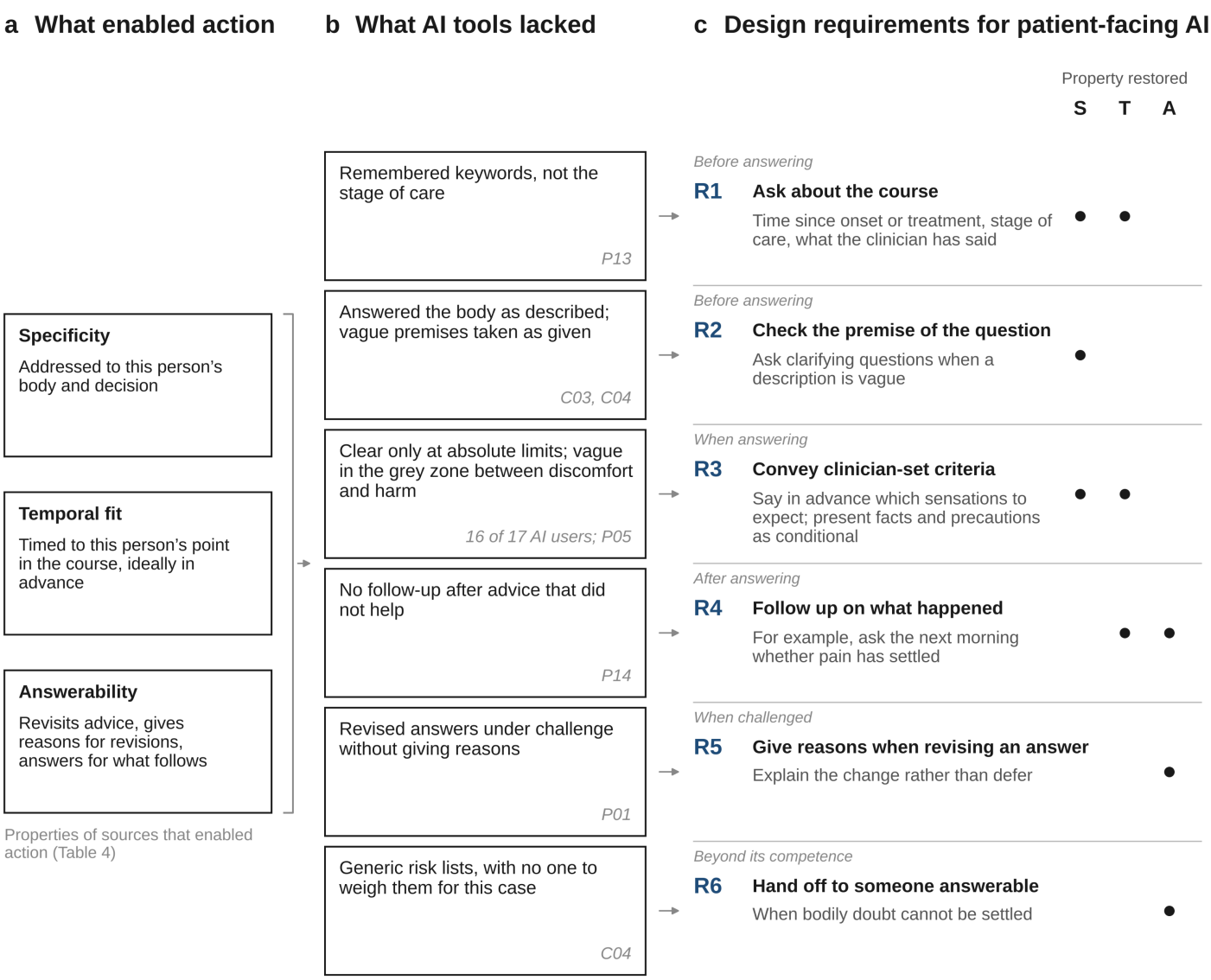


Figure 2

From what enabled action to design requirements for patient-facing AI. a, Three properties that distinguished the sources that enabled action (Table 4). **b**, What the AI tools participants used lacked, with the participants or clinicians whose accounts illustrate each gap. **c**, Design requirements R1–R6, ordered by when they apply in an exchange and developed in the Discussion; dots mark the properties each requirement restores (S, specificity; T, temporal fit; A, answerability). Panels summarise participants' and clinicians' accounts.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf