

Trust is fragile in AI-enabled healthcare

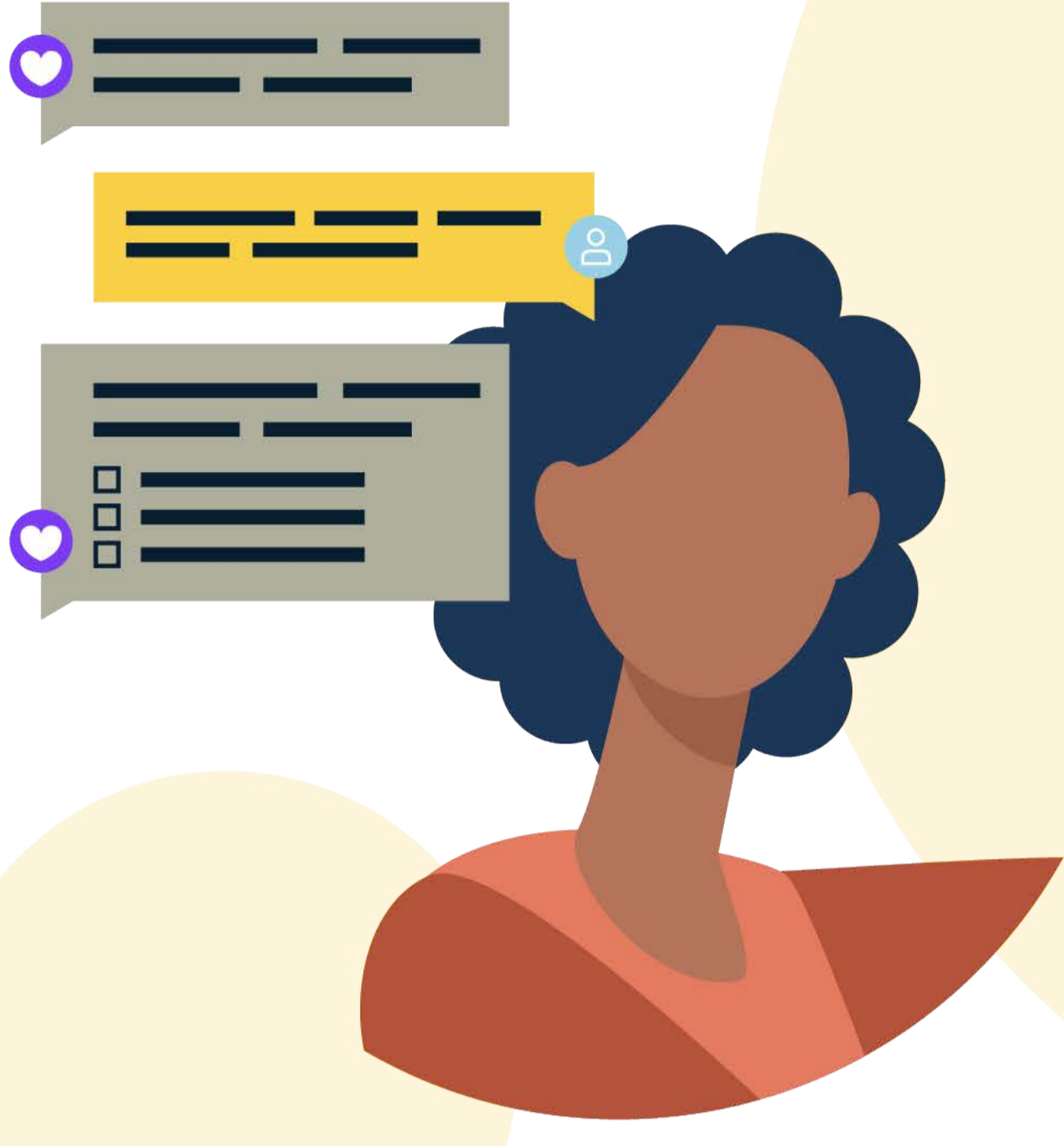
The intent of AI-powered solutions is to reduce user friction. When the output is accurate, people and the solution perform as intended. However, a challenge arises when the output is wrong and users don't question it. This is called **cognitive surrender**.

The problem

- Users defer to AI recommendations
- Personal judgment is bypassed
- AI becomes the decision-maker

When this happens in health, the consequences aren't theoretical — they're real.

Case example: Babylon Health



In 2019, Babylon Health launched an AI symptom checker used by millions and adopted by the UK's National Health Service.

Users often followed its direction on seeking HCP advice without questioning it, treating the system as a decision maker. When it labeled symptoms "low risk," some delayed seeking medical care; when it advised seeing a doctor, they complied automatically.

With a review by the Royal College of GPs in the UK flagged cases of misdiagnosis, trust didn't fade, it vanished overnight. By 2023, Babylon Health had gone bankrupt, highlighting how trust in AI can displace human judgment in health decisions.

The trust paradox

The insight: A new decision model

The traditional thinking model:

System 1: fast & instinctive

System 2: slow & deliberate

Introducing the newest addition to the model:

System 3: artificial thinking

In System 3, users don't evaluate — they adopt without questioning. The AI output becomes "their" decision.

The more a user trusts a platform, the less they question it.

In patient-facing solutions, that works well for:

- Small lifestyle changes
- Building habits
- Everyday wellness choices

But it creates risk for:

- Symptom recognition or tracking
- Medication adherence tracking
- Making health decisions

Trust is not neutral — it changes behavior.

Don't build AI that replaces thinking Build AI that supports thinking

Frictionless guidance in lower-stakes, everyday care

In lower-stakes situations, communication should prioritize ease and momentum, helping individuals take simple, practical steps without overthinking.

The focus is on making actions feel accessible and low effort, supporting ongoing engagement with everyday wellness and habit-building.

This can be achieved through light, suggestive language such as "you might try..." or "some people find..." which offers guidance without pressure and keeps the experience intuitive and easy to follow.

Intentional engagement in higher-stakes moments

In situations where health decisions carry greater significance, communication should be designed to slow the pace slightly and support more thoughtful consideration. Use questions to encourage this: "Here's a pattern I'm seeing. What do you think is going on?" The goal is not to create barriers, but to ensure the individual remains actively involved in understanding what's happening and what it means for them.

This approach helps:

- Set realistic expectations around progress, outcomes, and potential side effects
- Build stronger alignment with the treatment approach, supporting adherence over time
- Enable individuals to feel more informed and confident in how their care is progressing
- Support a clearer understanding of whether a treatment plan feels appropriate for their needs

Practical steps can be taken to avoid cognitive surrender

Build patient trust through intentional engagement

Define which guidance and coaching categories are safe for frictionless delivery and which require user judgment

Frame AI as a partner not an authority

Design tone and syntax to promote questioning where needed

A trust audit — if users trust this completely, what's the worst possible outcome?

Talk to S3 Connected Health about applying behavioral science to create safer, more effective digital health solutions and communications.

You can reach us at: Info@S3ConnectedHealth.com