A NAMED FRAMEWORK · AI LITERACY FOR PRIMARY CARE

The Hardest Day Test

One standard for deciding whether a clinical AI tool is actually safe to use.

> **If a tool needs you at your best to be safe, it is not safe.**

1

Why the test exists

Every tool is chosen on its best day and used on its worst one.

The demo runs on a rested clinician, an unhurried schedule, and a cooperative patient speaking clean English in complete sentences. That patient exists in the demo. Most of us have never met them.

Care happens somewhere else: in the last hour, behind schedule, with the interpreter running late and the inbox full. A tool that is only safe under demo conditions has not been made safe. It has had its risk moved onto the person using it.

2

Define your hardest day before you evaluate anything

Be specific. A vague standard fails no tool.

Write down your real conditions. For most of primary care, they look something like this:

You are forty minutes behind. The interpreter has not joined. A patient was added to your afternoon. Your inbox has forty items, and three of them are results you have not opened yet.

It is the last hour of clinic. You are hungry. You want to get home, maybe to pick up your child before child care closes, maybe because you told a friend you would be there, maybe because it has been a long day and you are tired.

None of that makes you a careless clinician. It makes you a person doing a demanding job under ordinary conditions. That day is not the exception. It is most Tuesdays.

That is the day the tool has to be safe on.

> **This cuts both ways, and that is the hopeful part.** A tool that stays safe on that day is usually the same tool that gives some of the day back to you. Safety and relief are not opposites here. They tend to arrive together, because both come from a tool that carries the load instead of handing it to you.
>
> The question was never whether a tool is impressive. It is whether it holds when you cannot.

3

The five questions

1. Does it still work when I am behind?
If safe use requires unhurried attention, it will not be used safely.

2. Does it fail loudly, or quietly?
Loud errors get caught. Quiet ones become the standard of care.

3. Does it need my full attention to be safe?
A tool that shifts vigilance onto the clinician has not reduced burden. It has renamed it.

4. Who does it work worst for, and how would I know?
Name the patients. Non-English speakers, code-switchers, low health literacy, hard of hearing, older adults. If you have no way to detect the failure, assume it is happening.

5. What happens the day it is down?
If there is no answer, the tool is not a tool. It is a dependency.

4

What passing and failing look like

A TOOL THAT PASSES	A TOOL THAT FAILS
· Safe when used quickly and imperfectly · Fails visibly, in a way you cannot miss · Degrades gracefully when you are distracted · You have watched it work on a messy, real, non-English visit · Care continues without it	· Safe only when you are careful and unhurried · Fails silently, in the parts you skim · Requires a new checking habit to stay safe · You have only ever seen it demoed · Nobody knows what happens when it is down

5

How to use it

In a demo. Ask the vendor to show you the tool failing. Not succeeding. Failing. What the failure looks like on screen, and how a busy clinician would notice.

In a pilot. Run it during your worst session of the week, not your best. A pilot scheduled for a light Tuesday morning tells you nothing.

In a procurement meeting. Say it out loud: *would this be safe on our hardest day?* It reframes the room faster than any list of features, because nobody in the room can argue that the hardest day is not real.

On yourself. If you are already using a tool and cannot answer question four, that is the work for this week.

Where this came from

I stopped asking whether a tool was impressive and started asking whether it held up on my hardest day, when I was behind and moving fast and not double-checking. Almost everything that looked good in a conference room failed that question.

The tools that passed were not the most advanced ones. They were the ones that stayed safe when I could not be careful.

Those tools exist. This is how I find them, and how I decide what to keep.

How to cite this

The Hardest Day Test. Magan G. *AI Literacy for Primary Care*, 2026. drgigimagan.com/resources

Use it in your clinic, your curriculum, your talk, or your procurement process. Please keep the name attached to it.

Dr. Gigi Magan, MD, MPH is a bilingual family physician and Certified Epic Physician Builder. Her work in safety-net primary care shapes how she evaluates every tool, and she writes and speaks about AI in primary care, digital health equity, and ambient documentation.

Read the newsletter: MedTech & Heart · drgigimagan.substack.com
Speaking and advisory: drgigimagan.com
More free tools: drgigimagan.com/resources