CHECKLIST · RESPONSIBLE AI

12 questions to ask before connecting AI to your organization's knowledge base.

For health, mental health and welfare organizations. Use these questions with your vendor, your IT team, or whoever is implementing the system. Based on the article "When the Knowledge Itself Is Poisoned" (in Hebrew) at dr-amir-tal.com.

01 / Information sources

- ☐ Which sources does the system draw on, and who approved each one?
- ☐ Is verified professional knowledge clearly separated from external content such as websites, articles or user feedback?
- ☐ If the system pulls content from the web, who reviews it before it enters the knowledge base?

02 / Permissions and updates

- ☐ Who is allowed to add, change or delete documents in the knowledge base?
- ☐ Is every change logged – who, what and when – and is anyone watching for unusual changes?

03 / Versions and currency

- ☐ What happens to an old protocol when a new one replaces it? Is it removed from the knowledge base?
- ☐ Is there a periodic review for outdated or contradictory documents?

04 / Transparency of recommendations

- ☐ Can clinicians see, alongside every significant recommendation, which sources it is based on?
- ☐ Can you reconstruct after the fact what a specific answer was based on?

05 / Testing and oversight

- ☐ Has the system been tested against deliberate attempts to mislead it (red teaming), not only routine scenarios?
- ☐ Has anyone checked that documents in the knowledge base contain no hidden instructions to the model (indirect prompt injection)?
- ☐ If the model was trained or fine-tuned on your data – on what data, and how was it checked for hidden "backdoors"?

> **An answer of "we don't know" to any of these questions isn't necessarily a reason to stop the project. But it is a reason to dig deeper.**