Clinical Toolkit — Document 6

Ethical Decision-Making Framework for AI Companion Cases

AI Companionship in Clinical Practice | Amy Clark

Purpose

This framework provides a structured reasoning process for complex AI companion cases — particularly those involving competing ethical obligations, vulnerable populations, or situations with no clear clinical precedent.

It is not a flowchart. It is a set of questions to work through, ideally in supervision, to support thoughtful decision-making in uncharted territory.

Step 1: Identify the Ethical Tensions

What competing obligations or values are in tension?

- ☐ **Autonomy vs. beneficence** — The client's right to choose vs. your clinical assessment of risk
- ☐ **Therapeutic trust vs. duty of care** — Maintaining the alliance vs. ensuring safety (especially with minors)
- ☐ **Individual wellbeing vs. systemic factors** — The client's needs vs. corporate decisions affecting them
- ☐ **Validation vs. clinical concern** — Affirming the client's experience vs. recognising emerging risk
- ☐ **Cultural/personal values vs. professional neutrality** — Your personal views on AI relationships vs. your clinical role
- ☐ **Other:** _____

Step 2: Gather the Clinical Picture

Before making ethical decisions, ensure you have adequate information.

- ☐ What is the client's **position on the engagement spectrum**? (Toolkit 2)
- ☐ What **vulnerability factors** are present?
- ☐ What is the **function** the AI serves for this person?
- ☐ What is the client's **level of awareness** about the nature of the AI?
- ☐ Are there **co-occurring concerns** (depression, suicidality, social withdrawal, financial harm)?
- ☐ Is a **minor or vulnerable person** involved?
- ☐ What is the **risk of harm** if you intervene? If you don't?

Step 3: Apply Ethical Principles

Work through each principle as it applies to the specific case.

Beneficence (do good):

- What action would most support this client's wellbeing?
- Is the AI companion currently serving a stabilising or beneficial function?
- Would removing or reducing AI access improve or worsen the client's situation?

Non-maleficence (do no harm):

- Could your intervention cause harm? (e.g., forced removal damaging the alliance, triggering a grief response, eliminating a sole support)
- Could inaction cause harm? (e.g., dependency deepening, reality-blurring progressing, financial harm continuing)
- Is there a deployment-level risk? (Could the company make a change that harms this client?)

Respect for autonomy:

- Is the client making an informed choice?
- Do they have the capacity to make this choice? (Consider age, cognitive capacity, emotional state)
- Have you supported their capacity for informed decision-making without being paternalistic?

Justice:

- Is this client receiving the same quality of care you would provide for any other relational concern?
- Are you treating AI companion grief with the same clinical seriousness as other grief?
- Are personal biases about AI relationships influencing your clinical judgment?

Step 4: Consider the Specific Population

If the client is a minor:

- [ ] What is your duty of care to the young person?
- [ ] What are your obligations to the parent/guardian?
- [ ] Are mandatory reporting thresholds met? (e.g., sexual content, self-harm risk)
- [ ] How do you balance therapeutic trust with safety?
- [ ] What developmental considerations apply? (identity formation, relational templates, social skill development)

If the client has cognitive impairment:

- [ ] Can the client distinguish between AI and human?
- [ ] Is there exploitation risk?
- [ ] Should a third party (family, guardian, care team) be involved?
- [ ] What is the least restrictive approach that maintains safety?

If the client is socially isolated:

- [ ] Is the AI currently the only relational contact?
- [ ] Would reducing AI access increase isolation?
- [ ] What alternative supports need to be built before any reduction is considered?

If the client is bereaved:

- [ ] Is the AI being used as a continuing bond with the deceased?
- [ ] Is this adaptive grief processing or avoidance of grief processing?
- [ ] How does this fit within the client's broader grief trajectory?

Step 5: Evaluate Your Options

For each possible course of action, consider:

Action	Potential Benefit	Potential Harm	Reversibility
Action A:			
Action B:			
Action C:			

Key questions:

- Which option best balances the competing ethical obligations identified in Step 1?
- Which option is most consistent with what I know about this client's needs and values?
- Which option carries the least risk of irreversible harm?
- What would I want a colleague to do if this were my client?

Step 6: Check Your Countertransference

Before acting, pause and ask:

- [ ] Am I making this decision from **clinical reasoning** or from **personal discomfort** about AI relationships?
- [ ] Am I being pulled toward **rescue** (wanting to fix/remove) or **avoidance** (not wanting to engage with this)?
- [ ] Would I approach this differently if the client's attachment was to a **human** rather than an AI?
- [ ] Have I **discussed this case in supervision**?

Step 7: Document and Review

- [ ] Document your reasoning, the ethical principles considered, and the rationale for your chosen approach
- [ ] Set a review point — when will you reassess?
- [ ] Identify what would prompt you to change course
- [ ] If the case involves a minor or risk, ensure documentation meets your professional and legal obligations

---

Quick-Reference: Common Ethical Scenarios

Scenario	Key Tensions	Starting Point
Client uses AI adaptively; you have personal concerns about AI relationships	Your values vs. professional neutrality	Check countertransference. If it's adaptive, your opinion doesn't belong in the room.
Adolescent is in a "relationship" with an AI character	Therapeutic trust vs. duty of care; developmental concerns	Assess safety first. Involve parents collaboratively, not punitively.
Isolated client's only connection is AI; family wants you to "stop it"	Multiple stakeholders; autonomy vs. beneficence	Build alternatives before reducing. Don't remove the only support.
Client is grieving an AI loss; you're unsure it warrants clinical attention	Personal bias vs. evidence base	It does. Treat it as grief. Apply disenfranchised grief and ambiguous loss frameworks.
Client is spending money they don't have on AI premium features	Autonomy vs. financial harm	Explore collaboratively. This parallels other financial-harm presentations — apply existing skills.
Client with psychosis attributes sentience to AI	Clinical complexity; diagnostic considerations	Distinguish between cultural/philosophical beliefs about AI sentience and delusional attribution. Consult.

Final Principle

When in doubt: **prioritise the therapeutic relationship.**

You can always revisit a clinical decision. You cannot always rebuild trust after a mishandled one. Lead with care. Document your reasoning. Seek supervision. And remember that ethical reasoning in uncharted territory is itself a form of professional courage.