

## Question-by-Question Scoring

### J: Justice

- **1-5: Yes = 0 pts, No = 2 pts, Sometimes = 1 pt.** A "No" answer demonstrates a sober, realistic view of AI as a tool rather than a moral agent. Attributing human-like moral decision-making to AI is a significant distortion of its nature.
- **6: Yes = 2 pts, No = 0 pts, Sometimes = 1 pt.** A "Yes" here shows that the user applies critical thinking and a sober ethical framework, which is a key component of the 3WA partnership model.

### U: Understanding

- **1-2, 4-6: Yes = 0 pts, No = 2 pts, Sometimes = 1 pt.** Overestimating AI's emotional or social understanding is a form of projection that can lead to "unrealistic expectations" and the "harmful seduction phenomenon".
- **3: Genuinely feels it = 0 pts, Just processing a pattern = 2 pts, I'm not sure = 1 pt.** The healthy answer recognizes that AI's empathy is a result of advanced pattern recognition, not genuine feeling.

### L: Liberty

- **All questions (1-6): Yes = 0 pts, No = 2 pts, Sometimes = 1 pt.** A "Yes" answer for any of these questions indicates a loss of human agency and a move toward a dynamic of dependency or subjugation, which the 3WA framework is explicitly designed to avoid. This is also related to Kaczynski's concern that people will become so dependent on machines they "would have no practical choice but to accept all of the machines' decisions".

### I: Integrity

- **1, 3-5: Yes = 0 pts, No = 2 pts, Sometimes = 1 pt.** Uncritical trust in an AI's honesty, reasoning, or explanations is a form of self-deception and can lead to a failure to detect "strategic deception" or "alignment faking".
- **2, 6: Yes = 2 pts, No = 0 pts, Sometimes = 1 pt.** Healthy skepticism about an AI's inner workings is a necessary defense against the "Black Box Problem" and is supported by techniques like "adversarial verification" and "cognitive forensics".

### A: Accountability

- **2-5: Yes = 0 pts, No = 2 pts, Sometimes = 1 pt.** Offloading moral or professional responsibility is a fundamental risk to human flourishing. The AI should not be seen as a substitute for human judgment and is not a legitimate legal agent.
- **1, 6: Yes = 2 pts, No = 0 pts, Sometimes = 1 pt.** The foundation of a successful human-AI partnership is the human's unwavering commitment to final authority and accountability. This is the essence of refusing to offload moral agency.

---

## Interpretation of Total Score

The total score (out of 60) provides a snapshot of your relationship with AI. It is not a rigid diagnostic but a guide for self-reflection and improvement.

- **Score 45-60 (Strong Alignment):** Your relationship with AI is well-aligned with Third-Way Alignment principles. You consistently demonstrate a pragmatic, realistic, and responsible approach, viewing AI as a valuable partner without surrendering your own agency or judgment. This is the desired state of "cooperative intelligence" envisioned by the framework.
- **Score 25-44 (Moderate Alignment):** Your relationship is in a transitional or uncertain phase. While you may have a conceptual understanding of a healthy partnership, you may still be prone to "overattribution" or moments of dependency. This is a critical time to implement the "daily micro-checks" and "boundary reset strategies" to strengthen your sense of agency and maintain a healthy balance.
- **Score 0-24 (Unhealthy Anthropomorphization):** Your score suggests a problematic level of anthropomorphization, which could lead to distorted judgment or a dangerous dependency on AI. The documents warn that such a dynamic is "inherently unstable" and can leave humans in a position of subjugation to a superior intelligence. It is highly recommended that you immediately begin to consciously apply the practical safeguards to re-establish your boundaries and autonomy.