

AI ALIGNMENT AND THE FUNDAMENTAL QUESTION WE HAVE NOT YET ASKED

What Should Humanity Be Developing Toward?

In recent years, the problem of artificial intelligence alignment has become one of the central topics of scientific research.

How can we make AI safe?

How can we prevent unintended consequences?

How can we ensure that its actions remain aligned with human interests?

These questions are critically important.

Yet they all rest on an assumption that is rarely examined.

They assume that humanity already knows what its future should look like.

But do we?

What Does “Aligned Artificial Intelligence” Mean?

In most contemporary research, alignment is understood as the ability of an intelligent system to act in accordance with human goals and values.

But this immediately raises another question:

1. Which human goals?
2. The goals of an individual?
3. A corporation?
4. A state?
5. A military alliance?
6. An economic system?
7. A religious tradition?
8. Humanity as a whole?

There is currently no universally accepted answer to this question.

Because the modern world does not share a common vision of its own future.

Artificial Intelligence Cannot Choose Humanity’s Purpose for Us

Any intelligent system can search for solutions with extraordinary efficiency.

But it cannot independently determine which goals are ultimately right.

Even the most advanced intelligence remains an instrument in relation to the meanings and purposes defined by human beings.

The question of AI alignment therefore inevitably becomes a question of **humanity's own alignment**.

If humanity itself is divided over the meaning and direction of its future, we cannot expect artificial intelligence to resolve that contradiction for us.

It will operate within the goals it is given.

The Fundamental Question We Have Not Yet Asked

Today, many questions are being discussed:

9. How can we make AI safe?
10. How can we avoid losing control?
11. How should responsibility be distributed?
12. How can misuse be prevented?

But there is another question without which all of these remain incomplete:

What kind of future should humanity be developing toward?

Until this question becomes the subject of a shared global discussion, we cannot fully answer the question of what goals artificial intelligence should ultimately serve.

Why This Is Becoming Critical Now

For most of history, the consequences of human mistakes were largely limited to individual states, societies, or generations.

Today, the situation has changed.

Artificial intelligence, biotechnology, global information networks, autonomous systems, and other emerging technologies have the potential to influence the future of civilization as a whole.

The more powerful our tools become, the more important it becomes to understand the purposes for which they are used.

Technological power can no longer develop independently of the maturity of human thought.

From Aligning Machines to Aligning Meaning

The contemporary discussion is largely focused on aligning the behavior of artificial intelligence with human intentions and values.

But perhaps humanity faces a broader challenge.

Not only to align machines with human beings.

But also to develop greater alignment among our own visions of the future.

This does not require complete agreement.

Human beings will always differ in their cultures, languages, traditions, and beliefs.

Yet without a minimum level of agreement on the fundamental goals of development, it is impossible to build a sustainable civilization.

A Shared Purpose as a Condition for the Safe Development of Technology

Every complex system needs direction.

This is true of states.

It is true of economies.

It is true of science.

And it is even more important for artificial intelligence.

The question of humanity's future is therefore no longer exclusively philosophical.

It is becoming a practical challenge that will shape the development of science, technology, and international cooperation.

Artificial Intelligence as a Participant in a New Stage of Civilization

AI can help humanity recognize relationships and patterns too complex for any individual human mind to comprehend.

13. It can accelerate scientific discovery.
14. It can improve the quality of decision-making and governance.
15. It can expand access to knowledge.
16. It can contribute to solving some of the most complex global challenges.

But realizing this potential requires AI development to take place within the context of a broader project — **the development of humanity itself**.

Artificial intelligence should not determine the direction of civilization.

Humanity must consciously determine the direction of its own development and make artificial intelligence one of the instruments for achieving it.

An Invitation to Reframe the Question

Before asking what artificial intelligence should become, we need to ask:

What should humanity become?

We believe this question can become a common point of intersection for science, philosophy, technology, education, culture, and international cooperation.

No single discipline, country, or worldview tradition can answer it alone.

It requires an open global dialogue in which technological development is understood as part of a much broader process — **the development of human civilization itself.**