

Some Thoughts on AI for 2026

Issues and 2026 predictions on AI pragmatic trends and areas of immediate needs for attention

By [Usama Fayyad](#)

December 2025

These thoughts are meant to capture what I believe is happening in the AI market and space. The idea is to align with areas requiring immediate actions. I use these observations on AI and my predictions for the biggest trends of 2026 to recommend alignment of our AI Strategy, differentiation, and our chance to be leaders in important aspects of applied AI (see companion doc on AI strategy alignment).

Introductory Background – controversy vs achievability

It is now established that AI is a force of major economic impact (especially in the knowledge economy). For background, the two top topics of interest are AGI (Artificial General Intelligence) and Physical AI (the intersection of AI with our physical world and environments and its use in physical [as opposed to digital] settings).

What is not controversial?

- Economic relevance of AI to our current knowledge economy
- The nature of jobs and information leverage will change fundamentally
- On narrow well-defined tasks, machine performance by far exceeds human performance
- The fact that “super-intelligence” will follow from achieving AGI – this is a straightforward logical deduction

What is Controversial?

- Claims that LLMs (and what are now becoming GLMs [Gigantic Language Models]) are approaching AGI. Such claims remain grossly exaggerated for today’s deep-learning-only approaches. We are still far from addressing the foundational problems of replicating human intelligence (let alone exceeding it with super-intelligence).
- The core problems of reasoning, planning, understanding, generalizing, and translating learnings across problem solving domains remain as difficult as they were decades ago.
- Whether we should be working towards AGI as a goal. This meaningful and deep controversy has caused many to call for a “halt” on working on AI.
- Physical AI is imminently achievable (per claims from NVIDIA and Elon Musk and others). This is more difficult than proclaimed.

While achieving AGI is a very interesting academic and scientific challenge, there is no question that we should approach it with forethought and ethical concerns. Precedents of taking caution are instructive and show the possibility of responsible R&D:

- Societies and research communities duly recognized the existential danger of nuclear weapons technology and introduced many controls and precautions
- We are getting duly worried about quantum computing and its consequences should it materialize, and efforts focused on mitigating its outcomes should it be realized are underway.

Thus we need to think as deeply about **AI Safety** (inclusive of Responsible AI) and the ethics and guardrails needed around developing the science of AGI.

What about Physical AI? *The difficulty of Physical AI is greatly underestimated.*

Interpolations abound on generalizing from the anticipated success in AI acceleration of the knowledge economy (which is rife with mundane, repetitive, and robotic tasks) to claiming this will extend to the physical world (workplace and personal). These arguments ignore many challenges (engineering and otherwise) of dealing with the physical complex and unpredictable environments.

My 2026 AI “Predictions”

1. **We will see AI transformations in work acceleration happen at scale**, rapidly redefining work in almost all areas of the knowledge economy.
2. **Physical AI will be the area of major challenges for this year and the next decade.** 2026 will be the year of realization of the degree of difficulty of replicating successful AI applications in the digital world to the challenges of the physical world. Here, I include the difficulties of dealing with several scientific problems, e.g. understanding and predicting cell and system biology as an example. **Physical AI will see major advances in more controlled environments:** manufacturing, warehouses, and many narrow physical tasks under controlled conditions (with major economic value realized).
3. **The build-up of unsustainable “Technical Debt” created by LLM prompts, vibe coding, and RAG pipelines.** 2026 will witness an inflection: the mess that has been building up as AI custom solutions are developed will be recognized as serious liabilities in organizations. When programming languages first emerged, organizations rushed to leverage computers and, in the process, created substantial technical debt and “legacy” systems that are expensive to deal with. A similar rush is now happening with natural language “prompt engineering”, vibe coding, and the uncontrolled evolution of RAG (Retrieval Augmented Generation) pipelines to deal with the inability to retrain LLMs on new data. The messy developments lead to technical debt accumulating so much faster -- without having a new way of *software and systems engineering* to help us approach it systematically.
4. **The over-investment in hardware infrastructure for large (gigantic) LLMs will rapidly collapse.** Smaller, cheaper, more efficient models will emerge as serious contenders. The current frenzy of infrastructure spend will likely lead to “bail outs” of the large tech companies who are over-investing since the U.S. views AI as a must-win race. In its attempt to compete in the “AI race” China will likely flood the market with faster, cheaper models as it is pressured to participate in this race but without the huge infrastructure capital of the U.S. This will re-define AI on edge devices and will create much cheaper AI models.
5. **Machine Learning (ML) returns to GenAI** - The LLMs of the “frontier models” have become so large that, paradoxically, the ability to train, fine-tune, or even update their training data has become prohibitive (training cost and cost of testing that previously acquired skills are not atrophied by new learning). ML will return to GenAI with the emergence of SLMs and predictive AI as alternatives to these monolithic “know-it-all” gigantic models. More importantly, ML will help with capturing knowledge and feedback

from human workers interacting with AI workflows: We will see a rise in ML-driven knowledge graph construction.

6. **Data importance will loom larger than ever, yet the challenges of data collection and management will continue to be the hardest blockers.** Data today is more of a liability than an asset (due to incompleteness, lack of quality, and lack of management know-how), but we will see a huge awakening to the importance of data as AI users realize the only differentiation possible is through proprietary data sets. Dealing with unstructured data will become a necessity and not just the domain of the few AI-haves (the large AI companies). I believe the data challenges will be recognized but not solved.
7. **Predictions markets are here to stay and grow.** The “*wisdom of the crowds*” is an old topic and the fundamental understanding of their predictive power has been studied. However, data, scale, and AI have led to a plethora of prediction markets with monetization businesses leading to a looming gambling epidemic where you can bet on anything – along with the horrible effects of gambling on society and individuals.
8. **AI-slop and the pandemic of misinformation media will hit major awareness and lead to crises:** podcasts, videos, false news articles, false science, and misinformation websites at scale enabled by generative AI will create an overwhelming situation and will begin to cause major over-reactions. Public confidence will be shaken and unfortunately, we have no good corresponding cures or counter measures in today’s world.
9. **Major cybersecurity attacks powered by AI will take place.** Bad actors have the advantages of early adoption and application of the AI technology, and the harms on our digital world will be huge with major economic and security consequences.
10. **AI Safety will grow as a major priority and an active field of R&D, learning, and practice** - from its infancy to necessity (like vaccines during the COVID-19 pandemic).
11. **Agentic AI for workflow automation becomes real and scalable:** how to capture workflows? how to automate them? and how to learn from human worker feedback to agent workers that employ ML will become a mainstream set of activities in 2026 and well into the next 3 years. This will make 2026 the year of workflow automation with many consequences on workers, organizations, and academia.
12. **Major expenditures in AI for military applications:** already started and it seems like major defense budgets will be allocated to AI technology on both sides of both oceans.
13. **Knowledge leverage and information navigation will become a high R&D priority:** new methods for capturing and summarizing new knowledge will become a high priority for all organizations. A dramatic example of information overload extending beyond business and consumers is the world of scientific research and publication where it has become virtually impossible to keep up with new papers and findings, even within narrow fields. The problems, challenges and scientific method remain the same, the tools for doing research and science will have to change dramatically.