

The AI Emergency: When AI Can Act Autonomously, Who Is In Control? New Keynote Topic.

When AI Can 'Think', Learn, Talk and **ACT**, Who Is in Control?

AI is graduating from answering our questions to being our [AGENT](#), making decisions and taking action on our behalf, and becoming increasingly [autonomous](#). The declared goal of most [frontier AI Labs](#) is not just to build powerful tools but to build '[digital humans](#)'.

That changes the stakes fundamentally: a mistake by autonomous agents – whether resulting from [misalignment](#) or from human operators with criminal or nefarious intent – can spread globally before anyone understands what happened, while governments and businesses are still deciding who is accountable. Dario Amodei, CEO of Anthropic, recently set forth that AI could “surpass all but the most skilled humans at finding and exploiting software vulnerabilities”. This is either very good theatre (as in 'doom marketing') or an increasingly worrisome existential threat.

To be sure, this is just one of many major issues that the '[too much of a good thing](#)' – dilemma of AI is confronting us with, while at the same time it is also clear that we cannot just forego the benefits of AI. This is truly a [HellVen](#) (Heaven and Hell) challenge, in its purest form:)

Here is a short list of the key issues I am looking at daily (I will publish more details on a dedicated microsite soon):

Safety, security, control and alignment with human values are simply

not the top priority of most leading AI companies (no matter what country they operate in). Despite their [frequent statements to the contrary](#) and some clever doom-marketing-release-us-from-liability tactics, the top priority of the leading AI labs is to dominate the market and to launch 'category killer products'. These labs will not regulate themselves, that much is certain.

We are thus accelerating towards [AGI](#) (Artificial General Intelligence) & [RSI](#) (Recursive Self Improvement), fuelled by hundreds of Billion \$ provided by VCs and (soon) a series of Trillion \$ IPOs. *Anything* that may slow things down has – until just now (Sept 2026) – not been a real option; not in the industry, and definitely not in politics, either. All of this is creating exponential existential risks for humanity (watch Bill Gates's blip [here](#)).

Jobs, work and employment, and education. We are totally unprepared for a world where many if not [most routines will be automated](#) due to the effects of what I call '[The Intelligence Revolution](#)' i.e. the rise of AI, agenting and eventually, General / Super Intelligence. In fact, in less than 15 years, I believe it is very likely that work will no longer be the most important factor in our lives – and 'work' may no longer be the key economic trigger for most of us, either (see the debate about Universal Basic Capital, and beyond, [here](#) on my blog, and here, on [GerdFeed](#)). We must urgently prepare for a world '[beyond routines and explicit knowledge](#)'. Human-only jobs are the future, and learning must focus on making us more human, not less. *Watch [this](#) summary video.

Geopolitics, a new world order and the erosion of democracy: The current [U.S.-led wave of techno-optimism](#) combined with destructive and blatant unilateralism, extreme capitalism and overall authoritarian political ideology is making the dangers of unchecked AI dramatically worse, fuelling an [AI arms race](#) and potentially creating a Darwinian, zero-sum economy in which very few will win and most will lose. If democracy is

relentlessly weakened by AI tools and human-like agents that cleverly manipulate us – boosting mistrust while fuelling personal disengagement – even AGI-enabled dictatorship scenarios where democracy is regarded as an obstacle become thinkable.

Human abdication and irrelevance: the more intelligent machines become, and the more integrated they will be in terms of interfaces and access (or even [embodiment](#)), the more we will become dependent on them. In fact, many would argue that human '[super-stupidity](#)' will eventually be the result of the AI's increasing 'super-intelligence', leading to [cognitive outsourcing](#) (followed by 'cognitive surrender'), reductionism, [loss of human agency](#) and a world of 'synthetic everything' – essentially rendering humans 'useless' and irrelevant.

Most importantly, there is currently **no 'Mission Control'** for AI (no matter which exact aspect of AI we are looking at). The UN has launched many [laudable initiatives](#) and has recently gained traction on this topic at the [latest Security Council meeting](#), but due to the security council veto rights structure and the lack of truly global governance instruments the UN currently lacks the power to set forth – much less enforce – a binding treaty. Time for a change.

In this keynote (which is constantly updated and can be deeply customized to specific audiences, demographics and locations), I explore the growing gap between AI's capability and human or societal readiness, between [tools and telos](#), zooming in on what leaders can do here and now to close it. From meaningful human oversight and clear limits on deployment to international cooperation, **I make the case for moving fast and FIXING things.**

The AI Emergency is a call to take responsibility and act while we still have

the power to shape what comes next.