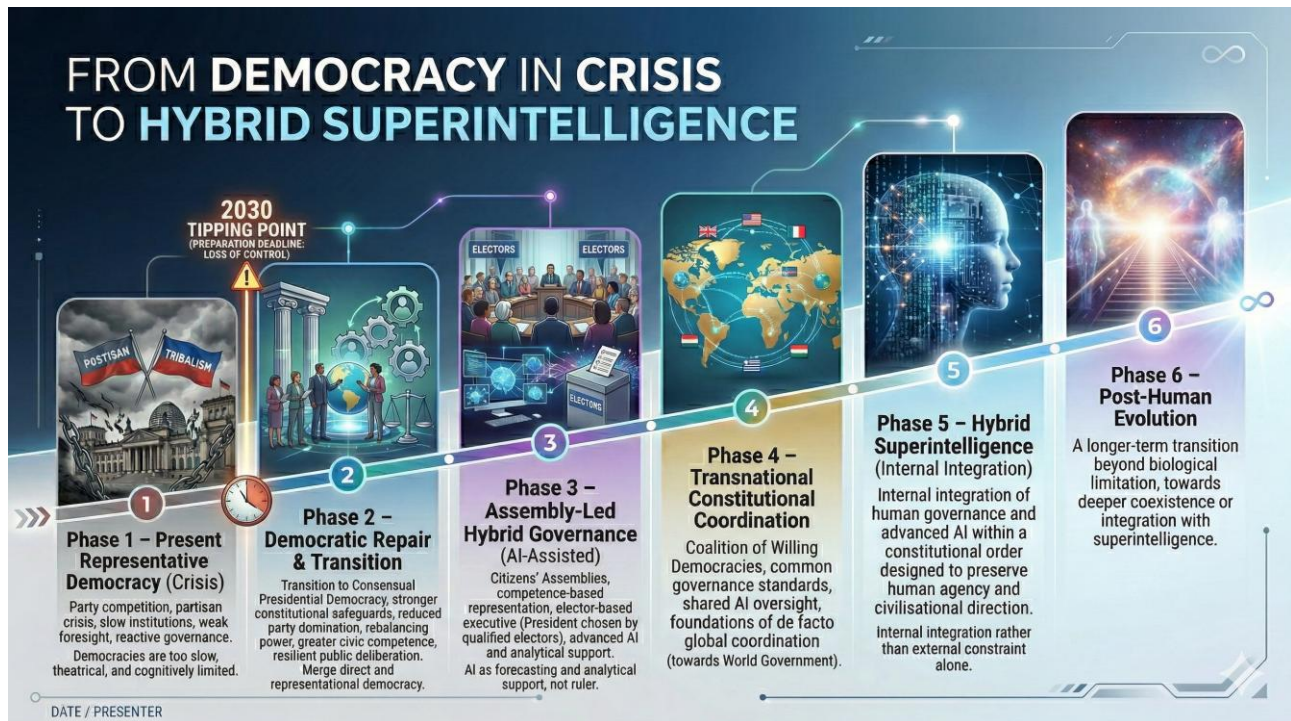


# Hybrid Superintelligence

Tony Czarnecki



## Introduction

I am sharing the first chapter of my pending book on “Hybrid Superintelligence” not just for you to read it but also to make critical comments. The reason why I am writing my seventh book on SAFE AI is that I no longer believe that humanity can safely coexist with Superintelligence in the present political system. For years I have argued that advanced AI may become the greatest existential threat humanity has ever created. What has changed is not the direction of my thinking but the speed with which events are moving. The world now changes so quickly that a manuscript can begin as a warning and end as an historical snapshot.

This new book is therefore not a fresh intellectual departure but an upgrade of a long line of thought. I began by putting the big question in my first book ‘Who could save humanity from superintelligence?’. I then moved on to the need for planetary coordination, democratic reform, political consensus, and eventually the transition towards a civilisation of transhumans. In ‘Prevail or Fail’ I assembled these strands into a larger architecture of civilisational shift. In this new work I am trying to take one further step. I want to show that the central political problem of the coming decade is not merely the regulation of AI tools. It is the redesign of governance itself.

My argument in this first part is simple but uncomfortable. Democracy, as we now practise it, is too slow, too partisan, too theatrical, too cognitively limited, and too poorly adapted to the exponential pace of technological change. I am sceptical that sufficiently powerful AI can be governed safely for long by external restraint alone. If AGI emerges before human governance and AI governance are both upgraded, humanity may lose control of its own future. That is why humanity must very rapidly begin redesigning its own governance.

For that reason, I want to begin with politics before I move to the subject of controlling and evolving with AI through a **Hybrid Superintelligence**. By this I mean a transitional form of civilisational co-governance, in

which human decision-making and advanced AI become internally integrated, so that advanced AI is guided from within the governing system, rather than merely constrained by humans from outside.

This transitional stage, if achieved, would also lay the foundation for humanity's eventual evolution from a biological species into a digital one, leading ultimately to a post-human phase.

When I wrote my earlier books, I still had the feeling that humanity possessed a little more time. The dangers were visible to me, but the world was still able to dismiss them as remote, speculative, or exaggerated. Today that dismissal is no longer intellectually serious. The release of powerful AI assistants, the speed of frontier model improvement, the acceleration of robotics, the rise of multimodal systems, and the increasing convergence of commercial, military, and geopolitical incentives have changed the context completely. The issue is no longer whether AI will transform civilisation. The issue is whether civilisation can transform itself fast enough to remain in control of AI.

I have for some time insisted that the decisive date is not the year when some philosopher is finally satisfied that a machine deserves the label of AGI, but the approximate point at which humans begin to lose effective control over the self-improvement and strategic behaviour of advanced AI systems. I have repeatedly argued that **2030 should be treated as the most likely tipping point for that loss of control** because political systems do nothing without a date or a threshold, e.g. 1.5C global temperature increase by 2030. That is why in an article *'Taking control over AI before it starts controlling us'*, London, October 2022, I said: "...we should take 2030, as the most likely date by which humans may lose an effective control over AI."

That sentence remains important because it turns a vague anxiety into a political horizon. Climate policy only began to acquire practical force once dates and thresholds were named. AI governance faces the same problem, but with even less time and even greater danger. If leaders continue speaking in loose terms such as 'in coming decades' or 'later this century', then action will always be postponed. That is why I continue to use 2030 as a public deadline for preparation, even if the actual year proves somewhat early or somewhat late.

There is another reason this book had to be written now. Much public discussion about AI safety remains technically interesting but politically shallow. It asks how to align models, how to reduce bias, how to improve monitoring, or how to regulate commercial deployment. These are all valid concerns, but they do not reach the root of the problem. If our governments themselves are intellectually, structurally, and morally unprepared for the age of AGI, then no amount of piecemeal regulation will be sufficient. A weak polity cannot govern a stronger intelligence for long.

This is why I now place democratic transition at the centre of the argument. In my earlier work, democratic reform was one of several indispensable principles of a safe civilisational shift. I still believe that. But I now see more sharply that political inadequacy is not merely one problem among many. It is the bottleneck through which all other solutions must pass.

# Chapter 1: Why Democracy Must Evolve Before AGI

## My Six-Book Journey Toward Hybrid Superintelligence

This new book is not an isolated reflection, but the latest stage in a line of thought I have developed over eight years through a series of books and articles. Each addressed a different part of the same civilisational problem. Together, they show both continuity in my concerns and an evolution in my proposed answers.

I began the journey with ‘[Who Could Save Humanity from Superintelligence?](#)’, published in 2018 after I had already been wrestling with the problem in 2017. In that first book I reviewed the major existential threats facing humanity and recognised that AI was different from most others. A global nuclear war or an artificial pandemic might devastate civilisation without necessarily eliminating every human being. Uncontrolled superintelligence, by contrast, could eventually make the entire species obsolete or extinct. That insight forced me to think beyond ordinary politics.

The next step was ‘[Federate to Survive!](#)’, published in 2020. There, I argued that humanity would need to act as a planetary civilisation if it wished to minimise existential risks. I came to see that no purely national approach would be enough. Yet I also recognised the practical difficulty of building a true world federation quickly enough.

Then came ‘[Democracy for a Human Federation](#)’ first issued in 2019 and later appearing in revised 2020 editions. In that work I turned to the question of how democracy itself would need to change if it were to support a safer and more coherent future. I proposed Consensual Presidential Democracy, combining elements of direct and representative democracy and trying to rebalance the power of governance between citizens and political elites.

In ‘[Becoming a Butterfly](#)’, published in 2021, I followed the argument further into human evolution. If superintelligence could not safely be controlled from outside forever, perhaps the only way forward would be for humans gradually to evolve with it, first into transhumans and eventually into posthumans. That book was an attempt to think seriously about the destiny of our species beyond biological limitation.

In ‘[2030 – Towards the Big Consensus](#)’, published in February 2023, I focused on the need for consensus as a political condition of civilisational survival. I argued that if humanity was to pass through the coming decades without tearing itself apart, then a new social contract would be required, one in which the voice of minorities could still be heard, and a more balanced democratic structure could emerge.

Finally, in ‘[Prevail or Fail](#)’, first published in 2023 and expanded in the January 2024 edition, I brought these strands together into the language of a civilisational shift. That book presented ten principles for coexistence with superintelligence, including democratic reform, AI development control, a de facto world government, a Global Welfare State, and a transition to transhuman governance.

*To summarize: I started with a question, which was also the title of my first book: ‘Who could save Humanity from Superintelligence?’ In that book, I have thoroughly reviewed key man-made existential risks to understand what might be done to mitigate them. Then, to answer fully the question I wrote five books, which were like the steps to solving a problem.*

What this new book adds is a clearer organising spine. It is no longer enough to say that humanity must federate, regulate AI, align values, and eventually evolve. I now want to show the transition pathway more concretely. The essential thesis is that governance must itself become augmented with AI as a hybrid entity before it is transformed into a **Hybrid Superintelligence**. The upgrading of political intelligence is not a side issue. It is the necessary bridge between our present condition and any safe coexistence with AI.

## Why Present Democracy Is Too Weak for the Age of AGI

Democracy remains one of humanity’s greatest achievements. It introduced legitimacy, accountability, peaceful alternation of power, and protection against arbitrary rule. I do not criticise democracy because I

prefer authoritarianism. I criticise complacency about democracy because the conditions under which it evolved no longer match the world now being created.

Today's representative democracies were designed for a slower world: a world of delayed communications, lower complexity, weaker interdependence, and a much slower pace of scientific and technological disruption. They depend heavily on political parties, periodic elections, media mediation, and civil services that can still assume that major changes will unfold over years rather than months. None of that assumption can safely be taken for granted now.

When a civilisation enters an era of exponential technologies, the first thing that becomes obsolete is linear governance. Yet most democracies still behave as though incremental reform, electoral theatre, and delayed committee reasoning are adequate responses to AI, synthetic biology, cyber conflict, or rapid geopolitical fragmentation. They are not.

I believe the weakness of present democracy can be described under four headings:

1. The first is **cognitive weakness**. The complexity of the modern world exceeds the unaided decision-making capacity of most political institutions. Ministers routinely decide on matters whose technical substance they barely understand.
2. The second is **structural weakness**. Party competition turns governance into continuous campaigning.
3. The third is **temporal weakness**. Democracies usually react late, often only when a crisis becomes visible to an average voter.
4. The fourth is **moral weakness**. We continue to assume that identical voting weight under all conditions is the sole meaning of democratic fairness, even when collective error may have species-level consequences.

Of these, temporal weakness may prove the most dangerous. In many policy areas delay is costly. In AI governance, delay may be fatal. By the time public opinion fully grasps what is happening, advanced systems may already have acquired strategic and operational roles from which they cannot easily be removed. That is why I have become increasingly sceptical of any theory which assumes that ordinary parliamentary inertia will somehow produce adequate safeguards in time.

There is an additional problem. Present democracies often reward performance over judgement. Leaders are selected for charisma, tribal appeal, rhetorical aggression, or media convenience rather than for wisdom, depth, foresight, or moral seriousness. This may be survivable in periods of relative stability. It is a dangerous luxury in an age when one erroneous decision about AI infrastructure, military integration, or geopolitical technological competition could shape the fate of billions.

For these reasons I no longer believe that preserving democracy means preserving its present procedures unchanged. If democracy is to survive, it must be redesigned.

## The Failure of Party Politics Under Exponential Change

The party system was historically useful. It aggregated interests, organised programmes, helped mobilise citizens, and created predictable routes to government. But under present conditions it has also become one of the chief obstacles to serious governance. Party competition pushes political life towards simplification, theatrical conflict, moral grandstanding, and short-term calculation. It turns complex civilisational questions into slogans that can be repeated within a media cycle.

This may be tolerable when the issues at stake are mainly symbolic. It becomes disastrous when the issues concern AI development control, biosecurity, deepfakes, frontier military systems, human augmentation, or the architecture of global governance. The danger is not merely that parties disagree. The danger is that they force disagreement into forms that are too shallow for the problems being faced.

In such an environment, truth is often subordinated to positioning. Good proposals may be rejected because they come from the wrong faction. Bad proposals may be embraced because they deliver immediate

advantage in the struggle for power. Long-term strategic thinking is crowded out by electoral timing. The result is a politics of permanent reaction to new events.

This is one reason I have for some time argued against governance by a single majority party. In my earlier democratic thinking, coalition government was already preferable because it reduced the domination of a single tribal machine and forced at least some degree of compromise. I still believe that. But the problem now appears deeper. The entire mechanism by which parties convert votes into executive control may itself be increasingly unfit for an age of civilisational peril.

I do not wish to abolish political association. People will always gather around values, priorities, and visions of the good society. But I propose to downgrade parties from executive gatekeepers into persuasive civic movements. They may continue to shape ideas, policies, and public discourse, but they should no longer monopolise the route to government.

## **Consensual Presidential Democracy Revisited**

At this point I need to revisit one of my own earlier proposals. In ‘Democracy for a Human Federation’ and again in ‘2030 – Towards the Big Consensus’, I argued for Consensual Presidential Democracy. The aim was to rebalance governance by combining representative democracy with direct democratic correction, strengthening consensus, and preventing the domination of society by a single party machine.

I do not abandon that idea. On the contrary, I now see it as an important transitional model. But I do think it needs to be revised in the light of AGI and the faster civilisational timetable we now face.

The original strength of Consensual Presidential Democracy was that it recognised two truths at once: first, that representative democracy on its own had become imbalanced; second, that mass direct democracy on its own could also be unstable, populist, or manipulable. The answer I proposed was a structure, in which a President, a more balanced legislature, and a Citizens’ Senate, could help stabilise governance while still preserving popular legitimacy. One of the key messages of my last book ‘*Prevail or Fail*’ (January 2024 edition) was: “*The starting point would be the removal of political and social imbalances in societies by merging direct and representational democracy into a new type of democracy – Consensual Presidential Democracy.*”

What now needs revision is the route by which the top executive is selected and legitimised. In the new hybrid age, I no longer think the President should necessarily be elected through a direct mass contest shaped by party media machines. That model may still look democratic, but in practice it can produce highly theatrical leadership precisely when the human civilization’s existence requires an entirely new approach.

I therefore now think of Consensual Presidential Democracy as a transitional constitutional framework rather than a single rigid design. It could still start with the rebalancing of representative democracy through Citizens’ Assemblies or a Citizens’ Senate and stronger consensus rules based on coalition governments. But it would very quickly shift the election of the President away from the current partisan contest towards a body of qualified electors.

The office of President would thus survive but rapidly evolve as new challenges and opportunities emerge. It might become the centre of a competence-oriented constitutional order.

## **Citizens’ Assemblies, Electors, and Competence-based Legitimacy**

The most controversial part of my current thinking concerns competence. I recognise that this immediately raises fears of elitism, exclusion, bias, and technocracy. Such dangers are real, and any proposal in this field must confront them directly, transparently, and with credible safeguards against abuse. Yet I believe it would be equally irresponsible, in an age of unprecedented technological power and civilisational risk, to act as though the question of civic and political competence does not matter.

In an age when voters may indirectly determine the future of AGI governance, biosecurity, or the constitutional architecture of the evolving civilization, it is no longer enough to assume that democratic

fairness means treating civic ignorance and civic competence as politically identical in every circumstance. I do not suggest disenfranchising anyone. But what I propose is a governance system in which civic responsibility and engagement are rewarded and publicly encouraged.

That is why I have proposed a competence-based democratic structure. Every adult citizen would retain the right to vote. But citizens willing to complete a transparent, public, learnable civic competence process would gain fuller political weight and greater eligibility for positions in Citizens' Assemblies. Those assemblies would become the core of governance at local, regional, national, and eventually global levels.

The key point is that the assembly is not merely an advisory panel. It becomes the principal deliberative institution through which the public selects its most competent representatives and through which the executive is formed. This inserts an additional layer of judgement between mass opinion and executive power without abolishing public legitimacy.

I continue to think that the first layer of competence assessment could remain factual and civic in nature. The second layer, for those seeking representative office, would need to test, among others, the candidates' judgement, practical reasoning, and social responsibility. I am more cautious than before about the use of AI in evaluating such judgement. AI may assist, summarise, pre-score, and flag patterns, but it should not be the unquestioned arbiter of political fitness.

The same reasoning leads me towards an elector-based presidency. Many democracies have already accepted indirect methods for selecting at least some public offices, on the understanding that legitimacy can flow through more than one institutional channel. In the model I envisage, once a national or transnational assembly structure is in place, the President and Vice-Presidents would be chosen not through party-driven mass contests, but by a qualified body of electors drawn from competence-based democratic institutions. This would not make the system less democratic in substance. It would redefine legitimacy at the highest level so that competence, judgement, and responsibility carry greater weight.

## **The President Without Parties**

I now come to what may be one of the most important constitutional refinements in the new book. If the executive is to be more competent, less tribal, and more stable under conditions of rapid technological transformation, then the head of the executive should not owe office because he has won the party combat. The President should stand above parties, not as a ceremonial abstraction but as a real constitutional principle.

In the earlier phase of transition, a national body of electors could be drawn from Citizens' Assemblies and other competence-qualified civic structures. In a later phase, when a coalition of capable democracies begins to govern together more tightly, there could be a larger transnational electoral college. In the most advanced phase, if transhuman governors emerge as part of a hybrid constitutional order, the top executive could be selected by a very large body (thousands) of cognitively enhanced electors or governors (by Brain-Computer-Interface devices).

What matters here is not the exact numerical formula but the logic. The top executive would derive legitimacy from a chain of competence-based representation rather than from a direct emotional contest mediated by parties and mass manipulation. The President would then be responsible not for embodying one faction's temporary triumph but for maintaining constitutional balance, civilisational direction, and executive coherence.

This also preserves something important from my earlier thinking about the future of democracy: the need for a visible centre of responsibility. Pure assembly rule may become diffuse. Pure presidentialism may become theatrical or authoritarian. A consensual presidency selected by competent electors could provide continuity, accountability, and symbolic unity without sliding back into tribal politics.

It is possible that some readers will object that this looks too distant from familiar democracy. I understand that reaction. But I would answer that the point is not to preserve familiar political system which has become

a theatre. The point is to preserve human self-government under radically altered conditions. If the older forms can no longer do that safely, then clinging to them becomes sentimentality rather than principle.

## **The First Transition: From Representative Democracy to Hybrid Governance**

Having set out the weaknesses of present democracy and the need for a revised constitutional design, I can now describe the first transition more clearly. It is not yet the transition to transhuman rule. It is the transition from today's party-dominated representative democracies to a more competence-based, assembly-led and AI-assisted constitutional order.

This transition would still be recognisably democratic. Citizens would vote. Rights would remain. Public accountability would remain. Open debate would remain. But the route from public opinion to executive authority would be more filtered by competence, more stabilised by assemblies, and more insulated from party tribalism. Government would be formed by assemblies making decisions based on their members' competence rather than just random selection. The President would be selected by electors rather than by parties. Parties would influence public thought but would lose their monopoly over executive formation.

At the same time, government would become progressively AI-assisted. I do not mean ruled by AI. I mean supported by it in forecasting, scenario modelling, coordination, and risk analysis. This is not an optional improvement. In the coming age no serious government will be able to function intelligently without advanced AI assistance. The only real question is whether such assistance will be plural, transparent, constrained, and ethically grounded.

This first transition is crucial because it prepares the ground for everything that follows. Without a better democratic architecture, there can be no credible process for selecting Universal Values of Humanity, the subject feeding an intensive debate on AI Alignment in the last few years. Neither can there be any trustworthy route to AI development control, no legitimate 'coalition of the willing', or as I would call it a de facto World Government, at a global level, and no safe progression towards hybrid governance. Political redesign therefore comes before a deeper fusion of human and machine intelligence.

That is why this chapter lays down the foundation for a logical extrapolation of the civilizational, and ultimately, human evolution. If humanity does not learn to upgrade its governance before AGI becomes effectively uncontrollable, then the dream of safe civilizational progress embedded in a hybrid Superintelligence, may never materialize. We will either drift into chaotic dependence on systems we do not govern, or lurch into emergency authoritarianism in the name of survival. I reject both outcomes. My hope is that humanity still has time to choose a harder but better path: a deliberate constitutional evolution leading eventually to safe coexistence with superintelligence.

The key argument presented in this chapter is that the greatest political mistake of the coming decade would be to treat AI as only a regulatory problem. It is a constitutional problem, a civilisational problem, and finally an evolutionary problem. The form of governance under which humanity entered the industrial and digital eras is unlikely to be sufficient for the age of AGI. If democracy is to endure, it must become deeper in legitimacy, stronger in competence, more protected from tribal capture, and more capable of working with advanced AI without surrendering to it.

Finally, I should add that this pathway does not depend on the assumption that every deeper form of human–AI integration, including advanced brain-computer interfaces, will arrive quickly or without major technical, biological, or philosophical obstacles. Serious difficulties may arise from reliability, safety, scalability, consciousness, identity, or the sheer limits of neural integration itself. I do not dismiss those concerns. On the contrary, they may prove decisive in shaping both the speed and the form of any transition towards hybrid superintelligence. Yet even if deeper fusion remains partial, delayed, or permanently constrained, the central argument of this chapter still holds: humanity must redesign governance so that advanced AI can be guided within a more intelligent constitutional order, rather than confronted only from outside by institutions that are already becoming inadequate.

In the following chapters, I will move from diagnosis to design in greater detail. I will examine how competence-based democracy, Citizens' Assemblies, a revised Consensual Presidential Democracy, and an elector-based executive could work in practice, what objections such a model must answer, and why the transition must remain constitutional for as long as possible. My conviction remains what it has long been: if humanity wishes to prevail rather than fail, it must begin by upgrading the intelligence of its own governance.