Some extracts from Dario Amodei's 2024 essay, "Machines of Loving Grace", followed by comments from Richard D North (11 January 2025)

The Dario extracts are in the order they appear in his essay

_____

Dario Amodei's title, "Machines of Loving Grace"

RDN comment: A happy half hour can be spent on Wiki looking into the background of the poem, "All Watched Over By Machines of Loving Grace", by Richard Brautigan. It matters that either his title is lightly ironic or it is a give-away as Dario's deep dreaminess. Do we seriously believe that computers will attain "loving grace"?

The 1960s San Francisco hippy poet was hanging out with computer talent at the time. There were giddy dreams of a cyber world where machines did better than humans. These are not redundant thoughts, but the hippy dream itself turned rather sour. So the take-home for a modern revival of this stuff is at least complex. The dive into the poem will soon bring one to Ayn Rand and Alan Greenspan – the latter the genius, one might sourly say, behind the techno-optimism which brought us the great financial crash of 2007.

_____

Dario: "I think and talk a lot about the risks of powerful AI. The company I'm the CEO of, Anthropic, does a lot of research on how to reduce these risks. Because of this, people sometimes draw the conclusion that I'm a pessimist or "doomer" who thinks AI will be mostly bad or dangerous. I don't think that at all. In fact, one of my main reasons for focusing on risks is that they're the only thing standing between us and what I see as a fundamentally positive future.

RDN comment: I am tempted to think the cynical thought. Dario's essay may be a diversionary or corrective attempt to place Anthropic at the heart of commercially viable AI (in biology, food tech, etc) in the face of its having been seen as a virtue-seeking or virtue-signalling pro-bono enterprise. This thought is sustained by the plain fact that when we come

(later in the piece) to Dario's loftier ambitions for AI's potential for human political development, he is much more modest. So this essay could be seen as a deft draft reconciliation of very present commercial possibilities and rather distant liberal democratic dreams.

Viewed more kindly, the first two-thirds of the piece are a straightforward account of the obvious merit of AI in technical research. The last, loftier sections about AI and human development look like wishful speculation: a commonplace variety of dreamy liberalism.

Dario: I can think of hundreds of scientific or even social problems where a large group of really smart people would drastically speed up progress, especially if they aren't limited to analysis and can make things happen in the real world (which our postulated country of geniuses can, including by directing or assisting teams of humans).

RDN comment: Isn't the problem that, say, 100 well-chosen clever people empowered (by whom?) could transition from thinking to action whilst radically mis-reading just how contested both the thinking and actions will be when put before unreconstructed human beings?

Dario: Biology is probably the area where scientific progress has the greatest potential to directly and unambiguously improve the quality of human life.

RDN comment: Yes. But even whilst we already see a fast track AI route to better health, that doesn't begin to answer the problem that long-lived people will still progress from an arse-wiped babyhood to an arse-wiped senility.

The solution may be that Dario's quest for human meaning will be answered by a massive new appetite for personal interactions amongst populations with Universal Income in their bank accounts but no clear work-purpose remaining.

Indeed, it may be that with AI having put human authenticity at a discount, persons will reinvent themselves face-to-face. Virtue-added

might replace value-added. Tending to other people may come to be a sort of defiance against technology, or merely a luxury conferred by it.

Dario: Beyond even curing disease, biological science can in principle improve the baseline quality of human health, by extending the healthy human lifespan, increasing control and freedom over our own biological processes, and addressing everyday problems that we currently think of as immutable parts of the human condition.

RDN comment: Just these freedoms have become the more controversial the more we see the need for highly nuanced, sceptical, actually brave, regulatory and rationing intelligence in health care providers. How else to care for the well-being of young, impetuous gender-changers? Tossing out airy mantras about "more choice" doesn't address this sort of stuff.

Dario: To summarize the above, my basic prediction is that AI-enabled biology and medicine will allow us to compress the progress that human biologists would have achieved over the next 50-100 years into 5-10 years. I'll refer to this as the "compressed 21st century": the idea that after powerful AI is developed, we will in a few years make all the progress in biology and medicine that we would have made in the whole 21st century.

RDN comment: One longs to see it. But it is reasonable to be nervous as to how the amount of doctoring and nursing required can be paid for quite as quickly as the medical advances can be churned out.

Dario posits a 5-10 year period of massive medical advance. Fixing the politics of paying for that will be a much longer affair.

Dario: [On poverty eradication] Nevertheless, I do see significant reasons for optimism. Diseases have been eradicated and many countries have gone from poor to rich, and it is clear that the decisions involved in these tasks exhibit high returns to intelligence (despite human

constraints and complexity). Therefore, AI can likely do them better than they are currently being done.

RDN comment: Many countries which have increased their wealth and even their populations' wellbeing have been quite or very authoritarian. What is wholly unknown now is where and when affluence will tend to produce responsive government. In the right hands, AI might conduce to greater government responsiveness, but it could easily be deployed toward greater control.

Dario: Can the developing world quickly catch up to the developed world, not just in health, but across the board economically? There is some precedent for this: in the final decades of the 20th century, several East Asian economies achieved sustained ~10% annual real GDP growth rates, allowing them to catch up with the developed world.

RDN comment: Two or three of Dario's Asian Tigers rather make the point of a comment of mine, above.

See: The Illiberal Logic of Mission-Directed Governance, by Bryan Cheang

---

Dario: I am more optimistic about within-country inequality especially in the developed world, for two reasons. First, markets function better in the developed world, and markets are typically good at bringing down the cost of high-value technologies over time[25].

RDN comment: It is surely not unduly cynical to say that at the rate at which modern technologies are monetized, plutocrats (I don't say, plutocracies) are being created quite quickly and that several of their empires do seem to be curiously fiscally and mentally hegemonic within their sectors.

Dario: The opt-out problem. One concern in both developed and developing world alike is people opting out of AI-enabled benefits

(similar to the anti-vaccine movement, or Luddite movements more generally).

RDN comment: One might reasonably both opt out of AI's increasing untruthing of social media and opt in to its health benefits. BTW: the social media untruthing problem for democracy is already very large, and AI will presumably offer to put its energized creative plagiarism into the hands of everyone, mad, bad or sad – or of the over-weeningly egotistical. It is not easy to see AI's energised plagiarism as a force for good.

Dario: On the international side, it seems very important that democracies have the upper hand on the world stage when powerful AI is created. AI-powered authoritarianism seems too terrible to contemplate, so democracies need to be able to set the terms by which powerful AI is brought into the world, both to avoid being overpowered by authoritarians and to prevent human rights abuses within authoritarian countries.

RDN comment: I risk suggesting that this is the core of Dario's naivity. Even if the West has better AI than Anne Applebaum's "Autarchy, Inc.", that doesn't at all mean that bad actors couldn't deploy primitive AI to great effect, at home and abroad. An analogy: bad actors could wreak almost as much damage with "dirty" bombs in suitcases as with inter-continental ballistic missiles. Besides, controlling information flows in individual countries surely begins with controlling the airwaves and handsets, and authoritarians seem quite good at that.

Dario: This [Western, benign] coalition would on one hand use AI to achieve robust military superiority (the stick) while at the same time offering to distribute the benefits of powerful AI (the carrot) to a wider and wider group of countries in exchange for supporting the coalition's strategy to promote democracy (this would be a bit analogous to "Atoms for Peace").

RDN comment: Was not Atoms for Peace a damp squib? Of course we have deployed atomic science in medicine, etc. But we still have a military

nuclear stand-off, and thank goodness, perhaps. Even a welcome de-escalation in the nuclear arms race has not affected the basic dynamic of MAD.

_____

Dario: If we can do all this, we will have a world in which democracies lead on the world stage and have the economic and military strength to avoid being undermined, conquered, or sabotaged by autocracies, and may be able to parlay their AI superiority into a durable advantage.

RDN comment: The West may become a beacon of good governance. Right now, Autarchy, Inc. seems to be pretty effective in not letting the bug spread to their territories.

Besides, Autarchy, Inc. is fairly happy to wage asymmetrical war against the West. It figures the West hasn't the stomach for a full-on Crusade against authoritarianism, even when it misbehaves in its own backyard.

It is important to note that it isn't Autarchy Inc.'s strength which lets it meddle in Western social media, etc: it is the weakness of many Western minds which gives it an open door.

_____

Dario: In particular, in this environment democratic governments can use their superior AI to win the information war: they can counter influence and propaganda operations by autocracies and may even be able to create a globally free information environment by providing channels of information and AI services in a way that autocracies lack the technical ability to block or monitor.

RDN comment: I am fairly sure that the "information war" will be won in the West when a taste for fact-based pragmatism returns. AI seems more likely to contribute to bendy post-modern relativism than to good sense.

Also: see the next comment.

_____

Dario: Second, there is a good chance free information really does undermine authoritarianism, as long as the authoritarians can't censor it. And uncensored AI can also bring individuals powerful tools for undermining repressive governments.

RDN comment: At home, the point of authoritarians is that they can control their citizenry's ability to find and act on information. There is no obvious mechanism whereby the West's "good AI" can subvert authoritarian bad actors. Meanwhile, "Autarchy, Inc." and its Bad AI – like present day Fake News – does not succeed against the West by its genius or technical power, but because we have a generation of Westerners who mop up this stuff.

Dario: I am not suggesting that we literally replace judges with AI systems, but the combination of impartiality with the ability to understand and process messy, real world situations feels like it should have some serious positive applications to law and justice. At the very least, such systems could work alongside humans as an aid to decision-making.

RDN comment: I see potential in AI working in parallel, say with juries or judges, to see how well its results calibrate with (improve on, are weaker than) the human.

Dario: In a similar vein, AI could be used to both aggregate opinions and drive consensus among citizens, resolving conflict, finding common ground, and seeking compromise. Some early ideas in this direction have been undertaken by the computational democracy project, including collaborations with Anthropic. A more informed and thoughtful citizenry would obviously strengthen democratic institutions.

RDN comment: As my previous comment, I do see merit in this.

However, "driving consensus" and "seeking compromise" is not really the heart of the liberal democracy Dario seeks or admires. Politics is more a matter of keeping many discordant human aspirations in play without either violence or consensus. It is a managed and lived dynamic tension.

For instance, there is no answer as to whether we should base society on competition or co-operation. Nor as to how much freedom we must sacrifice for there to be a sufficiency of order. Maybe the dilemma for AI will that it may see where rational consensus lies and be no more compelling to either the left or the right in society than centrist parties are now.

Besides: is Dario misreading the potential of AI when he stresses its capacity to find consensus? Or, a little differently, is he misreading the human appetite or capacity for consensus? Is Dario in the position of believing AI will be able to give us good reason for giving up human variety in tastes and opinion?

Dario assumes “a more informed and thoughtful citizenry” would be a Good Thing. This assumption may merely remind us that Parliaments are already stuffed with people who are more “informed and thoughtful” than their voters. Free elections are supposed to be how we calibrate and then elect them.

As to the quality of “informed and thoughtful citizenry”, universities are intended to produce such people. It is not cynical to remark that now universities process a substantial minority of our young citizens, they appear to have been engines for groupthink Theory. Our humanities graduates have not really proved to be an adornment to liberal democracy. Indeed, much of our modern dilemma is that the smugness of the educated has alienated the animal spirits of those who literally and figuratively do much of the heavy lifting in society.

_____

Dario: Having a very thoughtful and informed AI whose job is to give you everything you’re legally entitled to by the government in a way you can understand—and who also helps you comply with often confusing government rules—would be a big deal.

RDN Comment: Yes

Dario: Or perhaps humans will continue to be economically valuable after all, in some way not anticipated by the usual economic models.

RDN comment: I can imagine that if AI takes care of generating and even distributing material wealth, then humans could retreat – advance – into personal creativity and personal care (for ourselves and others). Future humans may even find a way out of our current dilemma that being free to think about ourselves (our personal development) too often leads to a neurotic self-absorption. However, the possession of human consciousness has never been easy for individuals and it may not get any more so, even if we are freed from much need to think about our physical wherewithal.

Dario: Through the varied topics above, I've tried to lay out a vision of a world that is both plausible if everything goes right with AI, and much better than the world today. I don't know if this world is realistic, and even if it is, it will not be achieved without a huge amount of effort and struggle by many brave and dedicated people. Everyone (including AI companies!) will need to do their part both to prevent risks and to fully realize the benefits.

RDN comment: I don't think Dario has addressed the way that no other technology has been able to tick the "if everything goes right" box. They have all been Pandora's boxes. Besides, it is tautologous to say that "if everything goes right", we will have a better world. And there is no inevitability that "a huge amount of effort and struggle by many brave and dedicated peoples" can produce the outcome he insists on hoping for. Indeed, it is precisely to degree to which the future of AI is not in human control that is the issue.

Splashing hope everywhere is not really a useful contribution to considering the dimensions of unpredictability which surround the technology. I imagine we will go ahead with it, all guns blazing, and do some tinkering with it in regulatory terms, perhaps to some effect. But one can't put genies back into bottles, and so far our historical failure to do so has mostly been to our advantage, at least a species level. (Provided, for instance, one wasn't a religious martyr, a French or Russian

aristocrat, a veteran of nuclear tests, or a WW1 warrior, and so on and on.)

_____

Dario: It is similarly intuitive that people should have autonomy and responsibility over their own lives and choices. These simple intuitions, if taken to their logical conclusion, lead eventually to rule of law, democracy, and Enlightenment values. If not inevitably, then at least as a statistical tendency, this is where humanity was already headed. AI simply offers an opportunity to get us there more quickly—to make the logic starker and the destination clearer.

RDN comment: At this present juncture the Whig History account of human development (an account I love) may be halted or stalled. AI is the latest iteration of various technological advances which more conduce to the problem than seem likely to solve it. Technologically, Dario may be a genius, and may be one of those whose names go down in history. Intellectually, Dario is a self-declared liberal dreamer. There is a chance that he represents the future of such dreaming (bucking quite recent but troubling developments), and will be a lauded for it. Give it a hundred years, and we may get a picture of how his reputation shapes up.

ends