

The Law of Intended Consequences: Designing Possible Futures

Natasha VITA-MORE
PhD Candidate, University of Plymouth
Faculty of Technology
School of Computing, Communications and Electronics,
Centre for Advanced Inquiry in the Interactive Arts

Abstract. If advocates of AGI face the future knowing what AGI requires, there is a greater possibility that they will achieve it. This paper suggests looking to futurology as a means to evaluate the critical intersection of AI and AGI, and intended consequences of AGI's future. The logic of intended consequences has a two-fold meaning. First is the convergent unyielding focus on basic requirements of AGI forecasting in building goals and watching for discontinuities. Methods include environmental scanning, trend analysis, and risk assessment. Second is the divergent counter-advocacy of challenging assumptions concerning AGI and hidden wildcards. Methods include scenario development and strategic planning, with an emphasis on applied critical thinking. This pairing of convergent and divergent foci breaks an epistemological bridge between theories of shared knowledge used by development teams and the phenomenon of hindsight bias in seeing past events as more predictable than they in fact were. Since AI suffered a long winter, taking it to the next level entails designing possible futures for AGI while recognizing tricky assumptions along the way.

Keywords. Futurology, intended consequences, design strategy, possible futures

Introduction

It is often advised that unforeseen and unintended consequences of our actions could turn promising ideas into sorrowful events. History warns us of such consequences with stories of plans gone badly, and perhaps one of the most noted modern story is the 1972 campaign to ban DDT (Dichloro-Diphenyl-Trichloroethane), the synthetic pesticide. The campaign warned society about the side effects of DDT on ecology and human health, potentially causing cancer. Environmentalists and other activists fought ardently to ban DDT. However, the banning of DDT triggered an unintended consequence—a large resurgence of malaria in developing countries, causing the deaths of millions of people.

No one knows the future. Strategists, futurists, and theorists all work at developing new techniques and coordinating methodologies to better assess the future in developing plans of action. Yet, the best we can do within our current capabilities is to accept the unexpected.

In doing so, I suggest what I call “intended consequences”—a plan of action that designates efforts to continually scan the environment for triggers of possible unexpected events. This practice is crucial in assessing the environment inside and around the field of AGI to locate possible discontinuities.

For example, AGI's advocates intend to develop a strategy for increasing the internal environment of AGI by increasing the number of researchers, the amount funding, and the magnitude of academic support. To evaluate intended consequences, we might consider possible triggers that could bring about negative consequences from employing such a strategy. Issues might concern how the field of AGI will overcome bias toward AI, and vice versa; ethical practices of AGI in developing smarter-than-human intelligence; the coupling AGI with the technological Singularity; downsides of collaborations with cognitive science and neuroscience; and feasible outcomes of efforts to build a field among individuals who are also competing forces. The strategies for dealing with these issues come in focus when discerning preferred and alternative futures for AGI, causing the least amount of risk and the greatest potential for success. Here, the key element of resolving issues and lowering risk in developing the future of AGI is the characteristic of flexibility to change course, if necessary, while safeguarding obtainable goals.

1. Looking to Futurology

This paper suggests looking to futurology as a means to evaluate the critical intersection of AI and AGI, designing alternative futures for AGI, and the discerning intended consequences of AGI's future. Futurology [1] is the study of the future as an interdisciplinary field and which incorporates a wide range of sources and models in identifying drivers of change for mapping out possible futures. The practice of futurology includes methods for identifying and researching trends, particularly in technology, and developing methodologies for theories based on carefully assessed observations of the internal environment of AGI as a field and the external environment of AGI in relation to outside opportunities and threats. Because these variables may change, tracking them is crucial in identifying what drives the opportunities and threats. It requires investigating multidisciplinary fields within the social, science, technology, economics, political domains.

One of the most advantageous methods for identifying drivers of change is through environmental scanning, the "acquisition and use of information about events, trends, and relationships in an organization's external and internal environment"[2]. Environmental scanning is an appropriate method for the field of AGI because of its systematic method of collecting and analyzing information that might impact AGI in the future from external forces outside the field.

External forces include social trends, positive and negative events that might affect the economy and productivity of AGI, and the behavior of other domains which could influence or be influenced by AGI. This information helps to establish possible parallels in other fields, and also helps to identify the competition in gaining a competitive advantage. This information also becomes a basic knowledge tool in developing a strategy for bringing about preferred futures or alternative futures for AGI. Further, environmental scanning makes an organization's leaders more informed about shifting social changes and uncertainty, and how these shifts could affect their organizations. It also helps researcher

identify numerous information sources, domains, time frames, contacts, and transdisciplinary relationships to AGI.

Another effective method when working with a yet-to-be developed product such as AGI is the method known as backcasting—working backwards from a given point in the future. Since the scenario is to work backwards from an outcome to the present condition, it is crucial to begin with a clear understanding of where you are now and a doable vision for where you want to be in the future. In that the backcasting method involves tracing events backwards, detailing what took place along the way to get from the present condition to the future vision can be a trying task, but the knowledge gained from this method is well worth it. Therefore, backcasting is a logical map which connects the dots of intended circumstances, including risks that could affect the outcome and cause necessary changes in strategy.

2. Logic of Intended Consequences

Having a focused intended outcome and awareness of unmeasured, unexpected events, which could affect those outcomes, is what I call the “logic of intended consequences.” Here, the immediate priority must be flexibility in (i) an acceptance that core values are subject to change and (ii) an agreement that all analysis is incomplete. The logic is a similar to the logic in performing a sport, such as downhill skiing. While the trail is defined by signage, whether it is a beginner, intermediate or double diamond expert run, the terrain along the way can shift at any moment and the skier must be constantly scanning the environment for any discontinuities and then, in a moment’s notice, adjust his behavior to adapt. Strategists argue for clear, clean strategies and commitment; however, no matter how clever or informed the strategist, in truth future is unknown. It needs the transdisciplinary efforts of teamwork with a strong leadership and vision and multidisciplinary methodologies.

3. Critical Intersection of AI and AGI, and AGI's Future

How can AGI successfully embrace *and* disassociate itself from AI? One way to explore AGI’s future is to take a pause at the intersection of AI and AGI. We know that environmental scanning is a method for recognizing converging factors in assessing the environment around AGI. We know that backcasting is a significant method for assessing diverging factors which could cause discontinuities in a focused plan. But what about the intersection where AI and AGI meet?

One strategy might be to assert that AI is beneficial as a task-oriented arm of AGI, and to invite AI advocates to play a core, practical role. Alternatively, another strategy might be for AGI advocates to shame the currents of AI as severely narrow by 21st century standards, and lacking skill and foresight to address its "AI Winter" [3] by having retreated. However, a pragmatic and possibly beneficial option for a strategy might be to promote the time line of AI, its successes and failures, and focus on how it lead to AGI and an

“alternative future” for AI and AGI [4].

Systems thinking is one tool which allows the user to change the variables/agents and their influencers by degrees within a given environment. One software tool for developing interactive diagrams is STELLA. Because STELLA is a dynamic visual interface, it allows you to apply different variables and to explore potential outcomes, especially when asking “what if” questions and then observing what happens. Knowing the modeling and simulated details and having them available for colleagues, researchers, executives, funding sources, academic institutions and students will set a precedent that the field of AGI. This sets a precedent that its practitioners are fully aware of the upside and downside of AI; fully aware of complete and missing theories on general intelligence and consciousness; cognizant of the collaborative advantages within the fields of neuroscience and cognitive science; and skillfully aware of applications of virtual environments. Here, the strategy includes a taxonomy for the field of AGI, a timeline of events, a 3D illustrations, synergistic mind maps, and dynamic visual simulations, which are all part of a media library supporting and encouraging AGI’s future.

4. Convergent Unyielding Focus

“Conventional wisdom would have it that a crisis is the most common trigger for change” [5]. But experience shows that a small portion of organizations in real need of change are actually facing a crisis. Most change is actually needed not because of crisis but because the organization is facing inconsistency in its planning. For example, its strategic initiatives are not delivering or meeting expectations.

Taking the approach of intended consequences would remind the leaders to return to the basic requirements for building plans based on what the product or organization needs. This reasoning brings into play the “why” factor [6]. The "why" factor is a lucid, unambiguous clarification of purpose. For example, can factors be brought together to identify AGI's purpose? Is the purpose greater than human intelligence, a problem solving intelligence to assist humanity in decision-making, a new consciousness, a defense against other Strong unfriendly AI, a means to meet the pending Singularity, or a future noosphere? The answers to these questions will frame AGI’s needs, the AGI advocates' wants, which together become the basic requirements (essentially, the desires infused with needs).

An arduous task for any group is to establish what the group needs to do to succeed. Everyone has an opinion and all possibilities could work. How do we decide which strategy is worth investing time and resources? In truth, countless and varied strategies can bring about similar results, positive or negative. Strategies are fueled with unyielding focus—what you want and what you need, and then pulling together from different directions as much information as possible to move it along. The method of environmental scanning is the most practical and insightful tool for developing ideas, theories and practices which will help to steer the direction for AGI by discerning its needs and its advocates' wants, which are the basic requirements for planning its future.

5. Divergent Counter-Advocacy Focus

Divergent counter-advocacy requires moving away from the accepted goals in order to develop as objective as possible assessment of possible hidden triggers of fallout and other wildcards. Its purpose would be to challenge assumptions concerning AGI's future. Methods include backcasting, scenario development, and strategic planning, with an emphasis on applied critical thinking.

How can we connect the dots from a preferred outcome of AGI back to the present condition where there is limited research, funding and advocacy? If the assumption that humanity is not prepared to handle the concept of a super intelligence, is there possibility an element of society that would support the development of AGI. Can this be tapped into?

It is far better to "be comprehensive" by considering "all reasonable alternative actions, including no action. Estimate the opportunities lost by abandoning a technology [such as AGI], and take into account the costs and risks of substituting other credible options." [7] Further, a team specifically formed to function as an intervener or neutral outsider to mediate innovative thinking and assumptions about risk could be part of such a comprehensive plan of counter advocacy. And lastly, the theory of the Proactionary Principle suggests, "prioritize and triage" to "prefer the measure with the highest expectation value by giving priority to more certain over less certain threats, to irreversible or persistent impacts over transient impacts, and to proposals that are more likely to be accomplished with the available resources." [8]

6. Designing Preferred Futures

One key advantage for the field of AGI is that "AGI's strengths, if they can achieve results, will be to provide a creative value and content—design, logic, and emotion—that enriches and advances whatever is being modeled. [Whereas] [t]he predictable behavior of agents can be defined in AI models; not desire, creativity, and vision, which create a goal that would drive and pull out attributes not otherwise uncovered" [9].

Developing key concepts for expanding numerous scenarios regarding the strengths of AGI will, in turn, strengthen the entire field of AGI. Preferred futures are not a happenstance occurrence. They are calculated approximations of what could happen through a focused application of the methods and tools of futurology in picturing where we want AGI to be in the coming years.

Even though the environment of technological innovation is unstable and in flux, this ought not to interfere with designing strategies for the marketability of AGI concepts and the viability of AGI practices. The methodology of discovery within known environments of technology can be just as exciting as in new environments of technology. What circumstances have been innovatively engineered with 20th century technology which could be better handled by AGI? How can AGI alleviate some of the world's most troubling problems? How can AGI-built environments, products, and services be conceptualized and marketed to assist in complex decision-making?

There is a wealth of ideas for building interactive, virtual landscapes, and

possibility for AGI to turn public attention toward productive problem-solving, educational projects, and entertainment-value projects.

Concluding Remarks

If advocates of the field of AGI face the future knowing what the practice of AGI requires, there is a greater possibility that preferred futures can be achieved. This paper has attempted to show why the field of futurology is a beneficial means to evaluate intended consequences of AGI's future. The author suggests the tool of environmental scanning as an appropriate method for tracking trends, drivers of change, and recognizing possible opportunities and threats. This practice reveals a converging environment in and around the field of AGI. The second tool, backcasting, is suggested as an appropriate method for establishing a strategy, and making note of possible divergent discontinuities that might arise and developing counter-advocacy in challenging assumptions about AGI's future. This method is crucial for developing strategies for the field of AGI. Together, the tools of environmental scanning and backcasting help to move beyond preconceived notions about how knowledge is often used by development teams in building strategies and the problems concerning hindsight bias. It is far better to invest time and effort in building tools for strategizing the future than looking in a rear-view-window.

The field of AGI has great promise. Designing its future can be exhilarating and offer stunning rewards. Now is the time to start strategizing to bring it about.

References

- [1] Futurology was coined by Ossip K. Flechtheim in mid-1940s.
- [2] Choo, C. W. (1998) *Information Management for the Intelligent Organization: the Art of Scanning the Environment*. Medford, NJ: Information Today.
- [3] "AI Winter" first appeared in 1982 in reference to a downfall or collapse in the perception of artificial intelligence research and development.
- [4] Vita-More, N. (2006) "Wisdom [Meta-Knowledge] through AGI / Neural Macrosensing", Consciousness Reframed 8th International Research Conference, Plymouth University, UK.
- [5] Wheeler, S., McFarland, W., and Kleiner, A. (2008) "A Blueprint for Strategic Leadership", *Strategy + Business*, Winter 2007, pp. 44-55.
- [6] *Id.*
- [7] More, M. (2007) "Chapter 4", *Proactionary Principle* (unpublished manuscript).
- [8] *Id.*
- [9] Foltz, J. (2008) Email communication with Ms. Foltz of Alsek Research.
www.alsekresearch.com