Global Catastrophic Risk
INSTITUTE

http://gcrinstitute.org

A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy

Seth D. Baum
Global Catastrophic Risk Institute

Global Catastrophic Risk Institute Working Paper 17-1

Cite as: Seth D. Baum, 2017. A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy. Global Catastrophic Risk Institute Working Paper 17-1.

Note from the author: Comments welcome. Revisions expected. This version 12 November 2017.

Executive Summary

Artificial general intelligence (AGI) is AI that can reason across a wide range of domains. While most AI research and development (R&D) is on narrow AI, not AGI, there is some dedicated AGI R&D. If AGI is built, its impacts could be profound. Depending on how it is designed and used, it could either help solve the world's problems or cause catastrophe, possibly even human extinction.

This paper presents the first-ever survey of active AGI R&D projects for ethics, risk, and policy. The survey attempts to identify every active AGI R&D project and characterize them in terms of seven attributes:

- The type of institution the project is based in
- Whether the project publishes open-source code
- Whether the project has military connections
- The nation(s) that the project is based in
- The project's goals for its AGI
- The extent of the project's engagement with AGI safety issues
- The overall size of the project

To accomplish this, the survey uses openly published information as found in scholarly publications, project websites, popular media articles, and other websites, including 11 technical survey papers, 8 years of the *Journal of Artificial General Intelligence*, 7 years of AGI conference proceedings, 2 online lists of AGI projects, keyword searches in Google web search and Google Scholar, the author's prior knowledge, and additional literature and webpages identified via all of the above.

The survey identifies 45 AGI R&D projects spread across 30 countries in 6 continents, many of which are based in major corporations and academic institutions, and some of which are large and heavily funded. Many of the projects are interconnected via common personnel, common parent organizations, or project collaboration.

For each of the seven attributes, some major trends about AGI R&D projects are apparent:

- Most projects are in corporations or academic institutions.
- Most projects publish open-source code.
- Few projects have military connections.
- Most projects are based in the US, and almost all are in either the US or a US ally. The only projects that exist entirely outside US and its allies are in China or Russia, and these projects all have strong academic and/or Western ties.
- Most projects state goals oriented towards the benefit of humanity as a whole or towards advancing the frontiers of knowledge, which the paper refers to as "humanitarian" and "intellectualist" goals.
- Most projects are not active on AGI safety issues.
- Most projects are in the small-to-medium size range. The three largest projects are DeepMind (a London-based project of Google), the Human Brain Project (an academic project based in Lausanne, Switzerland), and OpenAI (a nonprofit based in San Fransisco).

Looking across multiple attributes, some additional trends are apparent:

- There is a cluster of academic projects that state goals of advancing knowledge (i.e., intellectualist) and are not active on safety.

- There is a cluster of corporate projects that state goals of benefiting humanity (i.e., humanitarian) and are active on safety.
- Most of the projects with military connections are US academic groups that receive military funding, including a sub-cluster within the academic-intellectualist-not active on safety cluster.
- All six China-based projects are small, though some are at large organizations with the resources to scale quickly.

Figure ES1 on the next page presents an overview of the data.

The data suggest the following conclusions:

Regarding ethics, the major trend is projects' split between stated goals of benefiting humanity and advancing knowledge, with the former coming largely from corporate projects and the latter from academic projects. While these are not the only goals that projects articulate, there appears to be a loose consensus for some combination of these goals.

Regarding risk, in particular the risk of AGI catastrophe, there is good news and bad news. The bad news is that most projects are not actively addressing AGI safety issues. Academic projects are especially absent on safety. Another area of concern is the potential for corporate projects to put profit ahead of safety and the public interest. The good news is that there is a lot of potential to get projects to cooperate on safety issues, thanks to the partial consensus on goals, the concentration of projects in the US and its allies, and the various interconnections between different projects.

Regarding policy, several conclusions can be made. First, the concentration of projects in the US and its allies could greatly facilitate the establishment of public policy for AGI. Second, the large number of academic projects suggests an important role for research policy, such as review boards to evaluate risky research. Third, the large number of corporate projects suggests a need for attention to the political economy of AGI R&D. For example, if AGI R&D brings companies near-term projects, then policy could be much more difficult. Finally, the large number of projects with open-source code presents another policy challenge by enabling AGI R&D to be done by anyone anywhere in the world.

This study has some limitations, meaning that the actual state of AGI R&D may differ from what is presented here. This is due to the fact that the survey is based exclusively on openly published information. It is possible that some AGI R&D projects were missed by this survey. Thus, the number of projects identified in this survey, 45, is a lower bound. Furthermore, it is possible that projects' actual attributes differ from those found in openly published information. For example, most corporate projects did not state the goal of profit, even though many presumably seek profit. Therefore, this study's results should not be assumed to necessarily reflect the actual current state of AGI R&D. That said, the study nonetheless provides the most thorough description yet of AGI R&D in terms of ethics, risk, and policy.

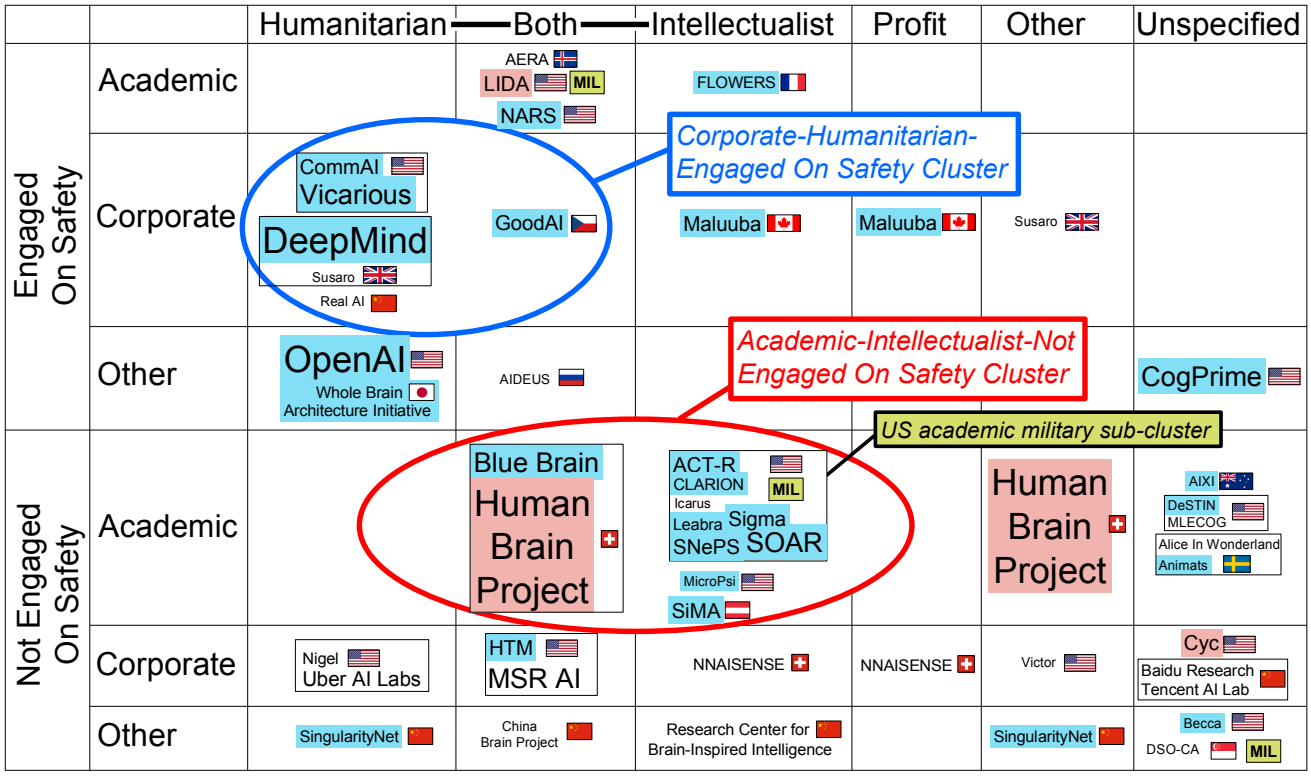

Figure ES1. Overview of the 45 identified AGI R&D projects characterized according to 7 attributes:

- **Institution type:** academic (mainly universities), corporate (public and private corporations), and other (government, nonprofit, and projects with no institution).
- **Open-source code:** blue background is code available open-source; red background is code available upon request.
- **Military connections:** projects labeled MIL have military connections.
- **Nationality:** Flags indicate the country that a project is based in. Australia, Austria, Austria, Canada, China, Czech Republic, France, Iceland, Japan, Russia, Singapore, Sweden, Switzerland, UK, USA
- **Stated goals:** humanitarian, both (humanitarian and intellectualist), intellectualist, profit, other (animal welfare, ecocentrism, and transhumanism), and unspecified (when project goals could not be identified). Projects that state multiple goals (aside from humanitarianism and intellectualism) appear multiple times in the figure.
- **Engagement on safety:** engaged projects either actively try to make their AGI safe or state support for other AGI safety work; not engaged projects either openly dismiss concerns about AGI safety or have no openly stated activity on safety.
- **Size:** font size indicates project size.

Acknowledgments

Tony Barrett, Robert de Neufville, Steven Umbrello, Ben Goertzel, Gordon Irlam, and Roman Yampolskiy provided feedback on an earlier draft. Matthijs Maas provided some research assistance. Any lingering errors or other shortcomings are the author's alone.

This work was made possible by a grant from the Gordon R. Irlam Charitable Foundation. The views in this paper are the author's alone and do not necessarily reflect the views of the Global Catastrophic Risk Institute or the Gordon R. Irlam Charitable Foundation.

Contents

1. Introduction

Artificial general intelligence (AGI) is AI that can reason across a wide range of domains. The human mind has general intelligence, but most AI does not. Thus, for example, DeepBlue can beat Kasparov at chess, and maybe at a few other basic tasks like multiplication, but it cannot beat him at anything else. AGI was a primary goal of the initial AI field and has long been considered its "grand dream" or "holy grail".[1] The technical difficulty of building AGI has led most of the field to focus on narrower, more immediately practical forms of AI, but some dedicated AGI research and development (R&D) continues.

AGI is also a profound societal concern, or at least it could be if it is built. AGI could complement human intellect, offering heightened capacity to solve the world's problems. Or, it could be used maliciously in a power play by whoever controls it. Or, humanity may fail to control it. Any AI can outsmart humans in some domains (e.g., multiplication); an AGI may outsmart humans in all domains. In that case, the outcome could depend on the AGI's goals: whether it pursues something positive for humanity, or the world, or itself, or something else entirely. Indeed, scholars of AGI sometimes propose that it could cause human extinction or similar catastrophe (see literature review below).

The high potential stakes of AGI raise questions of ethics, risk, and policy. Which AGI, if any, should be built? What is the risk of catastrophe if an AGI is built? What policy options are available to avoid AGI catastrophe and, to the extent that it is desired, enable safe and beneficial AGI? These are all questions under active investigation. However, the literature to date has tended to be theoretical and speculative, with little basis in the actual state of affairs in AGI. Given that AGI may be first built many years from now, some speculation is inevitable. But AGI R&D is happening right now. Information about the current R&D can guide current activities on ethics, risk, and policy, and it can provide some insight into what future R&D might look like.

This paper presents the first-ever survey of active AGI R&D projects in terms of ethics, risk, and policy. There have been several prior surveys of R&D on AGI and related technologies (Chong et al. 2007; Duch et al. 2008; Langley et al. 2008; de Garis et al. 2010; Goertzel et al. 2010; Samsonovich 2010; Taatgen and Anderson 2010; Thórisson and Helgasson 2012; Dong and Franklin 2014; Goertzel 2014; Kotseruba et al. 2016). However, these surveys are all on technical aspects, such as how the AGI itself is designed, what progress has been made on it, and how promising the various approaches are for achieving AGI. This paper presents and analyzes information of relevance to ethics, risk, and policy—for example, which political jurisdictions projects are located in and how engaged they are on AGI safety. Additionally, whereas the prior surveys focus on select noteworthy examples of AGI R&D projects, this paper attempts to document the entire set of projects. For ethics, risk, and policy, it is useful to have the full set—for example, to know which political jurisdictions to include in AGI public policy.

Section 1.1 explains terminology related to AGI. Section 2 reviews prior literature on AGI ethics, risk, and policy. Section 3 presents the research questions pursued in this study. Section 4 summarizes the methodology used for this paper's survey. Section 5 presents the main survey results about AGI R&D projects and other notable projects. Section 6 concludes. Appendix 1 presents the full survey of active AGI R&D projects. Appendix 2 presents notable projects that were considered for the survey but excluded for not meeting inclusion criteria.

1.1 Terminology

AGI is one of several terms used for advanced and potentially transformative future AI. The terms have slightly different meanings and it is worth briefly distinguishing between them.

[1] See e.g. Legg (2008, p.125); Pollock (2008); http://www.humanobs.org.

- *AGI* is specifically AI with a wide range of intelligence capabilities, including "the ability to achieve a variety of goals, and carry out a variety of tasks, in a variety of different contexts and environments" (Goertzel 2014, p.2). AGI is not necessarily advanced—an AI can be general without being highly sophisticated—though general intelligence requires a certain degree of sophistication and AGI is often presumed to be highly capable. AGI is also a dedicated field of study, with its own society (http://www.agi-society.org), journal (*Journal of Artificial General Intelligence*), and conference series (http://agi-conf.org).
- *Cognitive architecture* is the overall structure of an intelligent entity. One can speak of the cognitive architecture of the brains of humans or other animals, but the term is used mainly for theoretical and computational models of human and nonhuman animal cognition. Cognitive architectures can be narrow, focusing on specific cognitive processes such as attention or emotion (Kotseruba et al. 2016). However, they are often general, and thus their study overlaps with the study of AGI. *Biologically inspired cognitive architectures* is a dedicated field of study with its own society (http://bicasociety.org), journal (*Biologically Inspired Cognitive Architectures*), and conference series (http://bicasociety.org/meetings).
- *Brain emulations* are computational instantiations of biological brains. Brain emulations are sometimes classified as distinct from AGI (e.g., Barrett and Baum 2017a), but they are computational entities with general intelligence, and thus this paper treats them as a type of AGI.
- *Human-level AI* is AI with intelligence comparable to humans, or "human-level, reasonably human-like AGI" (Goertzel 2014, p.6). An important subtlety is that an AGI could be as advanced as humans, but with a rather different type of intelligence: it does not necessarily mimic human cognition. For example, AI chess programs will use brute force searches in some instances in which humans use intuition, yet the AI can still perform at or beyond the human level.
- *Superintelligence* is AI that significantly exceeds human intelligence. The term *ultraintelligence* has also been used in this context (Good 1965) but is less common. It is often proposed that superintelligence will come from an initial *seed AI* that undergoes *recursive-self improvement*, becoming successively smarter and smarter. The seed AI would not necessarily be AGI, but it is often presumed to be.

This paper focuses on AGI because the term is used heavily in R&D contexts and it is important for ethics, risk, and policy. Narrow cognitive architectures (and narrow AI) are less likely to have transformative consequences for the world. Human-level AI and superintelligence are more likely to have transformative consequences, but these terms are not common in R&D. Note that not every project included in this paper's survey explicitly identifies as AGI, but they all have the potential to be AGI and likewise to have transformative consequences. The survey does not exclude any projects that are explicitly trying to build human-level AI or superintelligence.

2. Prior Literature

2.1 Ethics

Perhaps the most basic question in AGI ethics is on whether to treat AGI as an intellectual pursuit or as something that could impact society and the world at large. In other words, is AGI R&D pursued to advance the forefront of knowledge or to benefit society? Often, these two goals are closely connected, as evidenced by the central role of science and technology in improving living conditions worldwide. However, they are not always connected and can sometimes be at odds. In particular for the present study, research into potentially dangerous new technologies can yield significant scientific and intellectual insights, yet end up being harmful to society.

Researchers across all fields of research often have strong intellectual values and motivations, and AGI is no exception. The question of whether to evaluate research in terms of intellectual merit or broader societal/global impacts is an ongoing point of contention across academia (Schienke et al. 2009). As with most fields, AI has traditionally emphasized intellectual merits, though there are calls for this to change (Baum 2017a). The intellectual pull of AGI can be particularly strong, given its status as a long-term "grand dream" or "holy grail", though the broader impacts can also have a strong pull, given the large potential stakes of AGI. In practical terms, an AGI project with intellectual motivations is more likely to view building AGI as a worthy goal in itself and to pay little attention to any potential dangers or other broader impacts, relative to a project motivated by broader impacts.

A second area of AGI ethics concerns what goals the AGI should be designed to pursue. This is the main focus of prior literature on AGI ethics. One line of thinking proposes "indirect normativity" or "coherent extrapolated volition", in which the AGI is designed to use its intellect to figure out what humanity wants it to do (Yudkowsky 2004; Muehlhauser and Helm 2012; Bostrom 2014). This proposal is motivated in part by concerns of procedural justice—everyone should have a say in the AGI's ethics, not just the AGI designers—and in part by concerns about the technical difficulty of programming the subtleties of human ethics directly into the AGI. Regardless, these proposals all call for the AGI to follow the ethics of *humanity*, and not of anything else.[2]

An alternative line of thinking proposes that the AGI should create new entities that are morally superior to humans. This thinking falls in the realm of "transhumanism" or "posthumanism"; AGI researchers de Garis (2005) and Goertzel (2010) use the term "cosmism". This view holds that AGI should benefit the cosmos as a whole, not just humanity, and proposes that the good of the cosmos may be advanced by morally superior beings produced by AGI. Whereas Goertzel (2010) stresses that "legacy humans" should be able to decide for themselves whether to continue in this new world, de Garis (2005) suggests that this world may be worth forming even if legacy humans would be eliminated. In contrast, Yampolskiy (2013) argues that AGI should only be built if they are expendable tools of benefit to their human creators.

Finally, there has also been some discussion of whether AGI should be built in the first place, though to date this has been a smaller focus of attention. Most discussions of AGI either support building it or do not seriously consider the matter because they presume its inevitability, as discussed by Totschnig (2017). Some arguments against building AGI are rooted in concern about catastrophe risk (e.g., Joy 2000); more on risk below. Others argue that even safe AGI should not be built. These include the fringe anarchist views of Kaczynski (1995, para. 174) and the more sober discussion of Totschnig (2017), though there has been much less outright opposition to AGI than there has been to similar transformative technologies like human enhancement.

2.2 Risk

The potential for AGI catastrophe is rooted in the notion that AGI could come to outsmart humanity, take control of the planet, and pursue whatever goals it is programmed with. Unless it is programmed with goals that are safe for humanity, or for whatever else it is that one cares about, catastrophe will result. Likewise, in order to avoid catastrophe, AGI R&D projects must take sufficient safety precautions.

Opinions vary on the size of this risk and the corresponding safety effort required. Some propose that it is fundamentally difficult to design an AGI with safe goals—that even seemingly minor mistakes could yield catastrophic results, and therefore AGI R&D projects should be very attentive to safety (Yudkowsky 2004; Muehlhauser and Helm 2012; Bostrom 2014). Others argue that an AGI can

[2] Baum (2017b) critiques coherent extrapolated volition proposals for focusing only on human ethics and not also on the ethical views that may be held by other biological species, by the AI itself, or by other entities.

be trained to have safe goals, and that this process is not exceptionally fragile, such that AGI R&D projects need to attend to safety, but not to an unusual extent (e.g., Goertzel and Pitt 2012; Bieger et al. 2015; Goertzel 2015; 2016; Steunebrink et al. 2016). Finally, some dismiss the risk entirely, either because AGI will not be able to outsmart humanity (e.g., Bringsjord 2012; McDermott 2012) or because it is too unlikely or too distant in the future to merit attention (e.g., Etzioni 2016; Stilgoe and Maynard 2017).

One common concern is that competing projects will race to launch AGI first, with potentially catastrophic consequences (Joy 2000; Shulman 2009; Dewey 2015; Armstrong et al. 2016). Desire to win the AGI race may be especially strong due to perceptions that AGI could be so powerful that it would lock in an extreme first-mover advantage. This creates a collective action problem: it is in the group's interest for each project to maintain a high safety standard, but it is each project's individual interest to skimp on safety in order to win the race. Armstrong et al. (2016) present game theoretic analysis of the AGI race scenario, finding that the risk increases if (a) there are more R&D projects, (b) the projects have stronger preference for their own AGI relative to others', making them less likely to invest in time-consuming safety measures, and (c) the projects have similar capability to build AGI, bringing them more relative advantage when they skimp on safety.

Barrett and Baum (2017a; 2017b) develop a risk model of catastrophe from AGI, looking specifically at AGI that recursively self-improves until it becomes superintelligent and gains control of the planet.[3] For this catastrophe to occur, six conditions must all hold: (1) superintelligence must be possible; (2) the initial ("seed") AI that starts self-improvement must be created; (3) there must be no successful containment of the self-improvement process or the resulting superintelligence, so that the superintelligence gains control of the planet; (4) humans must fail to make the AI's goals safe, such that accomplishment of the goals would avoid catastrophe; (5) the AI must not make its goals safe on its own, independent of human efforts to make its goals safe; and (6) the AI must not be deterred from pursuing its goals by humans, other AIs, or anything else. The total risk depends on the probability of each of these conditions holding. Likewise, risk management can seek to reduce the probability of conditions (2), (3), (4), and (6). Risk management is one aspect of AGI policy.

2.3 Policy

AGI policy can be understood broadly as all efforts to influence AGI R&D, which can include formal policies of governments and other institutions as well as informal policies of people interested in or concerned about AGI, including the researchers themselves. AGI policy can seek to, among other things, fund or otherwise support AGI R&D, encourage certain ethical views to be built into AGI, or reduce AGI risk. Sotala and Yampolskiy (2015) review a wide range of AGI policy ideas, focusing on risk management.

Much of the prior literature on AGI politics emphasizes the tension between (1) hypothetical AGI developers who want to proceed with inadequate regard for safety or ethics and (2) a community that is concerned about unsafe and unethical AGI and seeks ways to shift AGI R&D in safer and more ethical directions. Joy (2000) argues that the risk of catastrophe is too great and calls for a general abandonment of AGI R&D. Hibbard (2002) and Hughes (2007) instead call for regulatory regimes to avoid dangerous AGI without completely abandoning the technology. Yampolskiy and Fox (2013) propose review boards at research institutions to restrict AGI research that would be too dangerous. Baum (2017c) calls for attention to the social psychology of AGI R&D communities in order to ensure that safety and ethics measures succeed and to encourage AGI R&D communities to do more on their own.

[3] The model speaks in terms of AI in general, of which AGI is just one type, alongside other types of AI that could also recursively self-improve. This distinction is not crucial for the present paper.

One policy challenge comes from the fact that AGI could be developed anywhere in the world that can attract the research talent and assemble modest computing resources. Therefore, Wilson (2013) outlines an international treaty that could ensure that dangerous AGI work does not shift to unregulated countries. Scherer (2016) analyzes the potential for AGI regulation by the US government, noting the advantages of national regulation relative to sub-national regulation and suggesting that this could be a prelude to an international treaty. Goertzel (2009) analyzes prospects for AGI to be developed in China vs. in the West, finding that it could go either way depending on the relative importance of certain factors. Bostrom (2014) calls for international control over AGI R&D, possibly under the auspices of the United Nations. In order to identify rogue AGI R&D projects that may operate in secret, Hughes (2007), Shulman (2009), and Dewey (2015) propose global surveillance regimes; Goertzel (2012a) proposes that a limited AGI could conduct the surveillance.

Finally, prior literature has occasionally touched on the institutional context in which AGI R&D occurs. The Yampolskiy and Fox (2013) proposal for review boards focuses on universities, similar to the existing review boards for university human subjects research. Goertzel (2017a) expresses concern about AGI R&D at large corporations due to their tendency to concentrate global wealth and bias government policy in their favor; he argues instead for open-source AGI R&D. In contrast, Bostrom (2017) argues that open-source AGI R&D could be more dangerous by giving everyone access to the same code and thereby tightening the race to build AGI first. Shulman (2009) worries that nations will compete to build AGI in order to achieve "unchallenged economic and military dominance", and that the pursuit of AGI could be geopolitically destabilizing. Baum et al. (2011) query AGI experts on the relative merits of AGI R&D in corporations, open-source communities, and the US military, finding divergent views across experts, especially on open-source vs. military R&D.

It should also be noted that there has been some significant activity on AI from major governments. For example, the Chinese government recently announced a major initiative to become a global leader in AI within the next few decades (Webster et al. 2017). The Chinese initiative closely resembles—and may be derivative of—a series of reports on AI published by the US under President Obama (Kania 2017). Russian President Vladimir Putin recently spoke about the importance of AI, calling it "the future", noting "colossal opportunities, but also threats that are difficult to predict", and stating that "whoever becomes the leader in this sphere will become the ruler of the world" (RT 2017). But these various initiatives and pronouncements are not specifically about AGI, and appear to mainly refer to narrow AI. Some policy communities have even avoided associating with AGI, such as a series of events sponsored by the Obama White House in association with the above-mentioned reports (Conn 2016). Thus, high-level government interest in AI does not necessarily imply government involvement in AGI. One instance of high-level government interest in AGI is in the European Commission's large-scale support of the Human Brain Project, in hopes that a computer brain simulation could revive the European economy (Theil 2015).

3. Research Questions

The prior literature suggests several questions that could be informed by a survey of active AGI R&D projects:

How many AGI R&D projects are there? Armstrong et al. (2016) find that AGI risk increases if there are more R&D projects, making them less likely to cooperate on safety. Similarly, literature on collective action in other contexts often proposes that, under some circumstances, smaller groups can be more successful at cooperating, though large groups can be more successful in other circumstances (e.g., Yang et al. 2013).[4] Thus, it is worth simply knowing how many AGI R&D projects there are.

[4] The collective action literature specifically finds that smaller groups are often more successful at cooperating when close interactions reduce free-riding and the costs of transactions and compliance monitoring, while larger groups are often more

What types of institutions are the projects based in? Shulman (2009), Baum et al. (2011), Yampolskiy and Fox (2013), and Goertzel (2017a) suggest that certain institutional contexts could be more dangerous and that policy responses should be matched to projects' institutions. While the exact implications of institutional context are still under debate, it would help to see which institution types are hosting AGI R&D.

How much AGI R&D is open-source? Bostrom (2017) and Goertzel (2017a) offer contrasting perspectives on the merits of open-source AGI R&D. This is another debate still to be resolved, which meanwhile would benefit from data on the preponderance of open-source AGI R&D.

How much AGI R&D has military connections? Shulman (2009) proposes that nations may pursue AGI for military dominance. If true, this could have substantial geopolitical implications. While military R&D is often classified, it is worth seeing what military connections are present in publicly available data.

Where are AGI R&D projects located? Wilson (2013) argues for an international treaty to regulate global AGI R&D, while Scherer (2016) develops a regulatory proposal that is specific to the US. It is thus worth seeing which countries the R&D is located in.

What goals do projects have? Section 2.1 summarizes a range of ethical views corresponding to a variety of goals that AGI R&D projects could have. Additionally, Armstrong et al. (2016) finds that AGI risk increases if projects have stronger preference for their own AGI relative to others', which may tend to happen more when projects disagree on goals. Thus, it is worth identifying and comparing projects' goals.

How engaged are projects on safety issues? Section 2.2 reviews a range of views on the size of AGI risk and the difficulty of making AGI safe, and Section 2.3 summarizes policy literature that is based on the concern that AGI R&D projects may have inadequate safety procedures. Thus, data on how engaged projects are on safety could inform both the size of AGI risk and the policy responses that may be warranted.

How large are the projects? Larger projects may be more capable of building AGI. Additionally, Armstrong et al. (2016) find that AGI risk increases if projects have similar capability to build AGI. The Armstrong et al. (2016) analysis assumes that project capacity is distributed uniformly. It is worth seeing what the distribution of project sizes actually is and which projects are the largest.

Project capacity for building AGI is arguably more important than project size. However, project capacity is harder to assess with this paper's methodology of analyzing openly published statements. In addition to project size, project capacity could also depend on personnel talent, the availability of funding, computing power, or other resources, and on how well the project is managed. These factors are often not publicly reported. Another important factor is the viability of the technical approach that a project pursues, but this is not well understood and is a matter of disagreement among AGI experts. While it may be possible to assess project capacity with some degree of rigor, this paper's methodology is not suited for such a task, and thus it is left for future work. Instead, project size may be used as at least a rough proxy for project capacity, though caution is warranted here because it may be an imperfect or even misleading proxy.

4. Methodology

The paper's method consists of identifying AGI R&D projects and then describing them along several attributes. The identification and description was based on openly published information as found in scholarly publications, project websites, popular media articles, and other websites, with emphasis

successful at cooperating when cooperation benefits from larger total available resources (Yang et al. 2013). Thus, for example, one might want a small group for keeping a secret but a large group for fundraising for a fixed project.

placed on more authoritative publications. Identification and description were conducted primarily by the present author.[5]

In social science terminology, this methodology is known as the "coding" of qualitative data (Coffey and Atkinson 1996; Auerbach and Silverstein 2003). The data is qualitative in that it consists of text about AGI R&D projects; it is coded into quantitative form, such as "one academic project and three government projects". The coding scheme was initially developed based on prior literature and the present author's understanding of the topics and was updated during the coding process based on the present author's reading of the data (known as "in vivo" coding).

This methodology is fundamentally interpretive, rooted in the researcher's interpretation of the data. Some aspects of the data are not really a matter of interpretation—for example, the fact that the University of Southern California is an academic institution in the US. Other aspects are more open to interpretation. This includes which projects qualify as AGI R&D. Goertzel (2014, p.2) refers to the AGI community as a "fuzzy set"; this is an apt description. Different researchers may interpret the same data in different ways. They may also find different data as they search through the vast space of openly published information about AGI R&D. Thus, results should be read as one take on AGI R&D and not necessarily a true or complete reflection of the topic. Interested readers are invited to query the data for themselves and make their own interpretations. Appendix 1 contains full descriptions and explanations of coding judgments, citing the corresponding data. (Not all the data is cited—much of what was found is redundant or of limited relevance.)

4.1 Identification of AGI R&D Projects

AGI R&D candidate projects were identified via:

- The present author's prior knowledge.
- Keyword searches on the internet and in scholarship databases, mainly Google web search and Google Scholar.
- Previous survey papers (Chong et al. 2007; Duch et al. 2008; Langley et al. 2008; de Garis et al. 2010; Goertzel et al. 2010; Samsonovich 2010; Taatgen and Anderson 2010; Thórisson and Helgasson 2012; Dong and Franklin 2014; Goertzel 2014; Kotseruba et al. 2016).
- The entire available contents of *Journal of Artificial General Intelligence* (December 2009 through December 2016).[6]
- The proceedings of the most recent AGI conferences (2011 to 2017).
- Several online lists (http://www.agi-society.org/resources; http://bicasociety.org/cogarch/architectures.htm; http://realai.org/labs; http://2ai.org/landscape; https://en.wikipedia.org/wiki/Cognitive_architecture#Notable_examples).
- Feedback from colleagues listed in the Acknowledgments.
- Additional literature and webpages identified via all of the above.

Each identified project was put into one of three categories:

- Active AGI R&D projects (Appendix 1). These are projects that are working towards building AGI. The included projects either identify as AGI or conduct R&D to build something that is considered to be AGI, human-level intelligence, or superintelligence.

[5] Thanks go to Matthijs Maas for research assistance on characterizing some projects.

[6] Two additional journals, *Biologically Inspired Cognitive Architectures* and *Advances in Cognitive Systems*, were considered due to their AGI-related content, but they were not scanned in their entirety due to a lack of AGI projects reported in their articles.

- Other notable projects (Appendix 2). These include (1) inactive AGI R&D projects, defined as projects with no visible updates within the last three years; (2) projects that work on technical aspects of AGI but are not working towards building AGI, such as projects working on hardware or safety mechanisms that can be used for AGI; and (3) select narrow AI projects, such as AI groups at major technology companies.
- Other projects judged to be not worth including in this paper.

Only the active AGI R&D projects are included in the Section 5 data analysis. The other notable projects are reported to document related work, to clarify the present author's thinking about where the boundaries of AGI R&D lie, and to assist in the identification of any AGI R&D projects that have been overlooked by the present research.

Projects that only do R&D in deep learning and related techniques were excluded unless they explicitly identify as trying to build AGI. Deep learning already shows some generality (LeCun et al. 2015), and some people argue that deep learning could be extended into AGI (e.g., Real AI 2017). Others argue that deep learning, despite its remarkable ongoing successes, is fundamentally limited, and AGI requires other types of algorithms (e.g., Strannegård and Nizamani 2016; Wang and Li 2016; Marcus 2017). The recent explosion of work using deep learning renders it too difficult to survey using this paper's project-by-project methodology. Furthermore, if all of deep learning was included, it would dominate the results, yielding the unremarkable finding that there is a lot of active deep learning work. The deep learning projects that explicitly identify as trying to build AGI are much smaller in number, fitting comfortably with this paper's methodology and yielding more noteworthy insights.

4.2 Description of AGI R&D Projects

For each identified AGI R&D project, a general description was produced, along with classification in terms of the following attributes:

- *Type of institution*: The type of institution in which the project is based, such as academic or government.
- *Open-source*: Whether the project makes its source code openly available.
- *Military connections*: Whether the project has connections to any military activity.
- *Nationality*: The nation where the project is based. For multinational projects, a lead nation was specified, defined as the location of the project's administrative and/or operational leadership, and additional partner countries were tabulated separately.
- *Stated goal*: The project's stated goals for its AGI, defined as what the project aims to accomplish with its AGI and/or what goals it intends to program the AGI to pursue.
- *Engagement on safety*: The extent of the project's engagement with AGI safety issues.
- *Size*: The overall size of the project.

4.2.1 Type of Institution

The type of institution attribute has six categories:

- Academic: Institution conducts secondary education (e.g., colleges and universities).
- Government: Institution is situated within a local or national government (e.g., national laboratories). This category excludes public colleges and universities.
- Nonprofit: Institution is formally structured as a nonprofit and is not an academic institution (e.g., nonprofit research institutes).
- Private corporation: Institution is for-profit and does not issue public stock.

- Public corporation: Institution is for-profit and does issue public stock.
- None: Project is not based within any formal institution.

Some projects had two institution types; none had more than two. For the two-type projects, both types were recorded. Project participants were counted only if they are formally recognized on project websites or other key project documents. Some projects had more limited participation from many institutions, such as in co-authorship on publications. This more limited participation was not counted because it would make the entire exercise unwieldy due to the highly collaborative nature of many of the identified projects. This coding policy was maintained across all of the attributes, not just institution type.

4.2.2 Open-Source

The open-source attribute has three categories:

- Yes: Project has source code available for download online.
- Restricted: Project offers source code upon request.
- No: Project does not offer source code.

Projects were coded as yes if some source code related to their AGI work is open. Projects that have some, but not all, of their AGI code open are coded as yes. Projects that only have other, non-AGI code open are coded as no. Exactly which code is related to AGI is a matter of interpretation, and different coders may produce somewhat different results. The three open-source categories are mutually exclusive: each project is coded for one of the categories.

4.2.3 Military Connections

The military connections attribute has three categories:

- Yes: Project has identifiable military connections.
- No: Project is found to have no military connections.
- Unspecified: No determination could be made on the presence of military connections.

Military connections were identified via keyword searches on project websites and the internet at large, as well as via acknowledgments sections in recent publications. Projects were coded as having military connections if they are based in a military organization, if they receive military funding, or for other military collaborations. Projects were coded as having no military connections if they state that they do not collaborate with militaries or if the entre project could be scanned for connections. The latter was only viable for certain smaller projects. Unless a definitive coding judgment could be made, projects were coded as unspecified.

4.2.4 Nationality

The nationality attribute has two categories:

- Lead country: The country in which the project's administrative and/or operational leadership is based
- Partner countries: Other countries contributing to the project

One lead country was specified for each project; some projects had multiple partner countries.

4.2.5 Stated Goals

The stated goals attribute has six categories:

- Animal welfare: AGI is built to benefit nonhuman animals.
- Ecocentrism: AGI is built to benefit natural ecosystems.
- Humanitarianism: AGI is built to benefit humanity as a whole. This category includes statements about using AGI to solve general human problems such as poverty and disease.
- Intellectualism: AGI is built for intellectual purposes, which includes the intellectual accomplishment of the AGI itself and using the AGI to pursue intellectual goals.
- Profit: AGI is built to make money for its builders.
- Transhumanism: AGI is built to benefit advanced biological and/or computation beings, potentially including the AGI itself.
- Unspecified: Available sources were insufficient to make a coding judgment.

Some categories of goals found in prior AGI literature did not appear in the data, including military advantage and the selfish benefit of AGI builders.

For the coding of stated goals, only explicit statements were considered, not the surrounding context. For example, most AGI R&D projects at for-profit companies did not explicitly state profit as a goal. These projects were not coded under "profit" even if it may be the case that they have profit as a goal.

4.2.6 Engagement on Safety

The engagement on safety attribute has four categories:

- Active: Projects have dedicated efforts to address AGI safety issues.
- Moderate: Projects acknowledge AGI safety issues but lack dedicated efforts to address them.
- Dismissive: Projects argue that AGI safety concerns are incorrect.
- Unspecified: Available sources were insufficient to make a coding judgment.

Each project was coded with one of these categories. In principle, a project can be both active and dismissive, actively working on AGI safety while dismissing concerns about it, though no projects were found to do this.

4.2.7 Size

Finally, the project size attribute has five categories:

- Small
- Medium-small
- Medium
- Medium-large
- Large
- Unspecified: Available information was insufficient to make a coding judgment.

This is simply a five-point scale for coding size. Formal size definitions are not used because projects show size in different ways, such as by listing personnel, publications, or AGI accomplishments.

5. Main Results

This section presents the main results of the survey. Full results are presented in Appendices 1-2. Figure ES1 in the Executive Summary presents an overview.

5.1 The Identified AGI R&D Projects

45 AGI R&D projects were identified. In alphabetical order, they are:

1. ACT-R, led by John Anderson of Carnegie Mellon University
2. AERA, led by Kristinn Thórisson of Reykjavik University
3. AIDEUS, led by Alexey Potapov of ITMO University and Sergey Rodionov of Aix Marseille Université
4. AIXI, led by Marcus Hutter of Australian National University
5. Alice in Wonderland, led by Claes Strannegård of Chalmers University of Technology
6. Animats, a small project recently initiated by researchers in Sweden, Switzerland, and the US
7. Baidu Research, an AI research group within Baidu
8. Becca, an open-source project led by Brandon Rohrer
9. Blue Brain, led by Henry Markram of École Polytechnique Fédérale de Lausanne
10. China Brain Project, led by Mu-Ming Poo of the Chinese Academy of Sciences
11. CLARION, led by Ron Sun of Rensselaer Polytechnic Institute
12. CogPrime, an open source project led by Ben Goertzel based in the US and with dedicated labs in Hong Kong and Addis Ababa
13. CommAI, a project of Facebook AI Research based in New York City and with offices in Menlo Park, California and Paris
14. Cyc, a project of Cycorp of Austin, Texas, began by Doug Lenat in 1984
15. DeepMind, a London-based AI company acquired by Google in 2014 for £400m ($650m)
16. DeSTIN, led by Itamar Arel of University of Tennessee
17. DSO-CA, led by Gee Wah Ng of DSO National Laboratories, which is Singapore's primary national defense research agency
18. FLOWERS, led by Pierre-Yves Oudeyer of Inria and David Filliat of Ensta ParisTech
19. GoodAI, an AI company based in Prague led by computer game entrepreneur Marek Rosa
20. HTM, a project of the AI company Numenta, based in Redwood City, California and led by Jeffrey Hawkins, founder of Palm Computing
21. Human Brain Project, a consortium of research institutions across Europe with $1 billion in funding from the European Commission
22. Icarus, led by Pat Langley of Stanford University
23. Leabra, led by Randall O'Reilly of University of Colorado
24. LIDA, led by Stan Franklin of University of Memphis
25. Maluuba, a company based in Montreal recently acquired by Microsoft
26. MicroPsi, led by Joscha Bach of Harvard University
27. Microsoft Research AI, a group at Microsoft announced in July 2017
28. MLECOG, led by Janusz Starzyk of Ohio University
29. NARS, led by Pei Wang of Temple University
30. Nigel, a project of Kimera, an AI company based in Portland, Oregon
31. NNAISENSE, an AI company based in Lugano, Switzerland and led by Jürgen Schmidhuber
32. OpenAI, a nonprofit AI research organization based in San Francisco and founded by several prominent technology investors who have pledged $1 billion

33. Real AI, an AI company based in Hong Kong and led by Jonathan Yan
34. Research Center for Brain-Inspired Intelligence (RCBII), a project of the Chinese Academy of Sciences
35. Sigma, led by Paul Rosenbloom of University of Southern California
36. SiMA, led by Dietmar Dietrich of Vienna University of Technology
37. SingularityNET, an open AI platform led by Ben Goertzel
38. SNePS, led by Stuart Shapiro at State University of New York at Buffalo
39. Soar, led by John Laird of University of Michigan and a spinoff company SoarTech
40. Susaro, an AI company based in the Cambridge, UK area and led by Richard Loosemore
41. Tencent AI Lab, the AI group of Tencent
42. Uber AI Labs, the AI research division of Uber
43. Vicarious, an AI company based in San Francisco
44. Victor, a project of 2AI, which is a subsidiary of Cifer Inc., a small US company
45. Whole Brain Architecture Initiative (WBAI), a nonprofit in Tokyo

Many of the projects are interconnected. For example, AIDEUS lead Alexey Potapov is an advisor to SingularityNET; Animats contributors include Claes Strannegård (Alice in Wonderland), Joscha Bach (MicroPsi), and Bas Steunebrink (NNAISENSE); parts of DeSTIN have been used in CogPrime; DeepMind and OpenAI collaborate on AGI safety research; CogPrime and SingularityNET are led by the same person (Ben Goertzel), and CogPrime technology is being used by SingularityNET; Blue Brain and the Human Brain Project were both initiated by the same person (Henry Markram) and share a research strategy; Maluuba and Microsoft Research AI have the same parent organization (Microsoft), as do the China Brain Project and RCBII (the Chinese Academy of Sciences); and ACT-R and Leabra were once connected in a project called SAL (an acronym for Synthesis of ACT-R and Leabra; see Appendix 2). This suggests an AGI community that is at least in part working together towards common goals, not competing against each other as is often assumed in the literature (Section 2.2).

5.2 Type of Institution

20 projects are based at least in part in academic institutions, 12 are at least in part in private corporations, 6 are in public corporations, 5 are at least in part in nonprofits, 4 are at least in part in government, and two (AIDEUS, Becca) have no formal institutional home. 4 projects are split across two different institution types: Animats and Soar are academic and private corporation, FLOWERS is academic and government, and DeSTIN is academic and nonprofit. Figure 1 summarizes the institution type data.

The preponderance of AGI R&D in academia and for-profit corporations is of consequence for AGI policy. The large number of academic projects suggests merit for the Yampolskiy and Fox (2013) proposal for research review boards. However, while academia has a long history of regulating risky research, this has mainly been for medical and social science research that could pose risk to human and animal research subjects, not for research in computer science and related fields. More generally, academic research ethics and regulation tends to focus on the procedure of research conduct, not the consequences that research can have for the world (Schienke et al. 2009). There are some exceptions, such as the recent pause on risky gain-of-function biomedical research (Lipsitch and Inglesby 2014), but these are the exceptions that prove the rule. Among academic AGI R&D projects, the trend can be seen, for example, in the ethics program of the Human Brain Project, which is focused on research procedure and not research consequences. Efforts to regulate risky AGI R&D in academia would need to overcome this broader tendency.

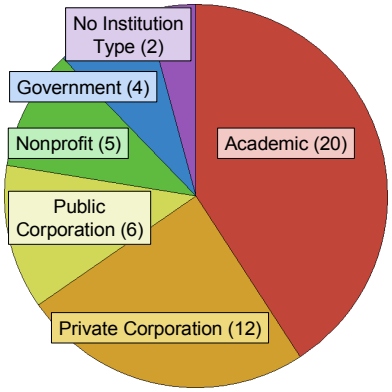

Figure 1. Summary of institution type data. The figure shows more than 45 data points because some projects have multiple institution types.

The preponderance of AGI R&D in academia is perhaps to be expected, given the longstanding role of academia in leading the forefront of AI and in leading long-term research in general. But the academic AGI community is rather different in character than the AGI risk literature's common assumption of AGI R&D projects competing against each other to build AGI first. The academic AGI community functions much like any other academic community, doing such things as attending conferences together and citing and discussing each others' work in publications in the same journals. The same can be said for the nonprofit and government projects, which are largely academic in character. Even some of the corporate projects are active in the academic AGI community or related academic communities. This further strengthens the notion of the AGI community as being at least partially collaborative, not competitive.

Among the for-profit corporations, two trends are apparent. One is a portion of corporations supporting long-term AGI R&D in a quasi-academic fashion, with limited regard for short-term profit or even any profit at all. For example, Cyc was started in 1984, making it among the longest-running AI projects in history. GoodAI, despite being structured as a for-profit, explicitly rejects profit as an ultimate goal. With less pressure for short-term profit, these groups may have more flexibility to pursue long-term R&D, including for safety mechanisms.

The other trend is of AGI projects delivering short-term profits for corporations while working towards long-term AGI goals. As Vicarious co-founder Scott Phoenix puts it, there are expectations of "plenty of value created in the interim" while working toward AGI (High 2016). This trend is especially apparent for the AGI groups within public corporations, several of which began as private companies that the public corporations paid good money to acquire (DeepMind/Google, Maluuba/Microsoft, and Uber AI Labs, which was formerly Geometric Intelligence). If this synergy between short-term profit and long-term AGI R&D proves robust, it could fuel an explosion of AGI R&D similar to what is already seen for deep learning.[7] This could well become the dominant factor in AGI R&D; it is of sufficient importance to be worth naming:

> *AGI profit-R&D synergy*: any circumstance in which long-term AGI R&D delivers short-term profits.

The AGI profit-R&D synergy is an important reason to distinguish between public and private corporations. Public corporations may face shareholder pressure to maximize short-term profits, pushing them to advance AGI R&D even if doing so poses long-term global risks. In contrast, private

[7] Similarly, Baum (2017a) observes synergies between groups interested in societal impacts of near-term and long-term AI, such that growing interest in near-term AI could be leveraged to advance policy and other efforts for improving outcomes of long-term AI, including AGI.

corporations are typically controlled by a narrower ownership group, possibly even a single person, who can choose to put safety and the public interest ahead of profits. Private corporation leadership would not necessarily do such a thing, but they may have more opportunity to do so. However, it is worth noting that two of the public corporations hosting AGI R&D, Facebook and Google, remain controlled by their founders: Mark Zuckerberg retains a majority of voting shares at Facebook (Heath 2017), while Larry Page and Sergey Brin retain a majority of voting shares at Google's parent company Alphabet (Ingram 2017). As long as they retain control, AGI R&D projects at Facebook and Google may be able to avoid the shareholder pressures that public corporations often face.

The corporate R&D raises other issues. Corporations may be more resistant of regulation than academic groups for several reasons. First, corporations often see regulation as a threat to their profits and push back accordingly. This holds even when regulation seeks to prevent global catastrophe, such as in the case of global warming (e.g., Oreskes and Conway 2010). Second, when the regulation specifically targets corporate R&D, they often express concern that it will violate their intellectual property and weaken their competitive advantage. Thus, for example, the US successfully resisted verification measures in the negotiation of the Biological Weapons Convention, largely out of concern for the intellectual property of its pharmaceutical industry (Lentzos 2011). To the extent that AGI corporations see AGI as important to their business model, they may resist regulations that they believe could interfere.

Finally, there are relatively few projects in government institutions. Only one project is based in a military/defense institution: DSO-CA. The other three government projects list intellectual goals in AI and cognitive science, with one also listing medical applications (Section 5.5). However, this understates the extent of government involvement in AGI R&D. Numerous other projects receive government funding, aimed at advancing medicine (e.g., Blue Brain, HBP), economic development (e.g., Baidu), and military technology (e.g., Cyc, Soar).

5.3 Open-Source

25 projects have source code openly available online. An additional three projects (Cyc, HBP, and LIDA) have code available upon request. For these 28 projects, the available code is not necessarily the project's entire corpus of code, at least for the latest version of the code, though in some cases it is. There were only 17 projects for which code could not be found online. For 3 of these 17 projects (Baidu Reseach, Tencent AI Lab, Uber AI Labs), their parent organization has some open-source code, but a scan of this code identified no AGI code. Figure 2 summarizes the open-source data.

The preponderance of open-source projects resembles a broader tendency towards openness across much of the coding community. Many of the projects post their code to github.com, a popular code repository. Even the corporate projects, which may have competitive advantage at stake, often make at least some of their code available.

Goertzel (2017a) distinguishes between two types of open-source projects: "classic" open-source, in which code development is done "out in the open", and "corporate" open-source, in which code development is conducted by project insiders in a closed environment and then released openly. Goertzel (2017a) cites OpenAI as an example of corporate open-source (note: OpenAI is a nonprofit project, not at a for-profit institution); his CogPrime would be an example of classic open-source. The classic/corporate distinction can matter for ethics, risk, and policy by affecting who is able to influence AGI goals and safety features. However, which open-source projects are classic or corporate is beyond the scope of this paper.

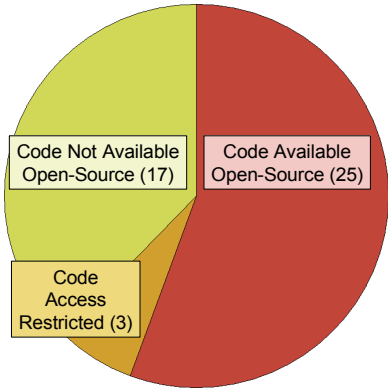

Figure 2. Summary of open-source data.

5.4 Military Connections

Nine projects have identifiable military connections. These include one project based in a military/defense institution (DSO-CA, at DSO National Laboratories, Singapore's primary national defense research agency) and eight projects that receive military funding (ACT-R, CLARION, Icarus, Leabra, LIDA, Sigma, SNePS, Soar). These eight projects are all based at least in part in US academic institutions. This follows the longstanding trend of military funding for computer science research in the US. Soar is also based in the private company SoarTech, which heavily advertises military applications on its website. Figure 3 summarizes the military connections data.

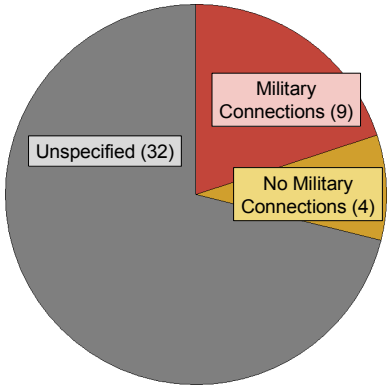

Figure 3. Summary of military connections data.

Only four projects were identified as having no military connections. Two of these (Aera, HBP) openly reject military connections; the other two (Alice in Wonderland, Animats) were sufficiently small and well-documented that the absence of military connections could be assessed. The other 32 projects were coded as unspecified. Many of these projects probably do not have military connections because they do not work on military applications and they are not at institutions (such as US universities) that tend to have military connections. Some projects are more likely to have military connections. For example, Microsoft has an extensive Military Affairs program that might (or might not) be connected to MSR AI.

The publicly available data suggest that the Singaporean and American militaries' interest in AGI is, by military standards, mundane and not anything in the direction of a quest for unchallenged military dominance. DSO-CA appears to be a small project of a largely academic nature; a recent paper shows DSO-CA applied to image captioning, using an example of a photograph of a family eating a meal (Ng et al. 2017). The project does not have the appearance of a major Singapore power play. Similarly, the military-funded US projects are also largely academic in character; for most of them,

one would not know their military connections except by searching websites and publications for acknowledgments of military funding. Only Soar, via SoarTech, publicizes military applications. The publicized applications are largely tactical, suggestive of incremental improvements in existing military capacity, not any sort of revolution in military affairs.

It remains possible that one or more militaries are pursuing AGI for more ambitious purposes, as prior literature has suspected (Shulman 2009). Perhaps such work is covert. However, this survey finds no evidence of anything to that effect.

5.5 Nationality

The projects are based in 14 countries. 23 projects are based in the US, 6 in China, 3 in Switzerland, 2 in each of Sweden and the UK, and 1 in each of Australia, Austria, Canada, Czech Republic, France, Iceland, Japan, Russia, and Singapore. Partner countries include the US (partner for 7 projects), the UK (4 projects), France and Israel (3 projects), Canada, Germany, Portugal, Spain, and Switzerland (2 projects), and Australia, Austria, Belgium, Brazil, China, Denmark, Ethiopia, Finland, Greece, Hungary, Italy, the Netherlands, Norway, Russia, Slovenia, Sweden, and Turkey (1 project). This makes for 30 total countries involved in AGI R&D projects. Figure 4 maps the nationality data.

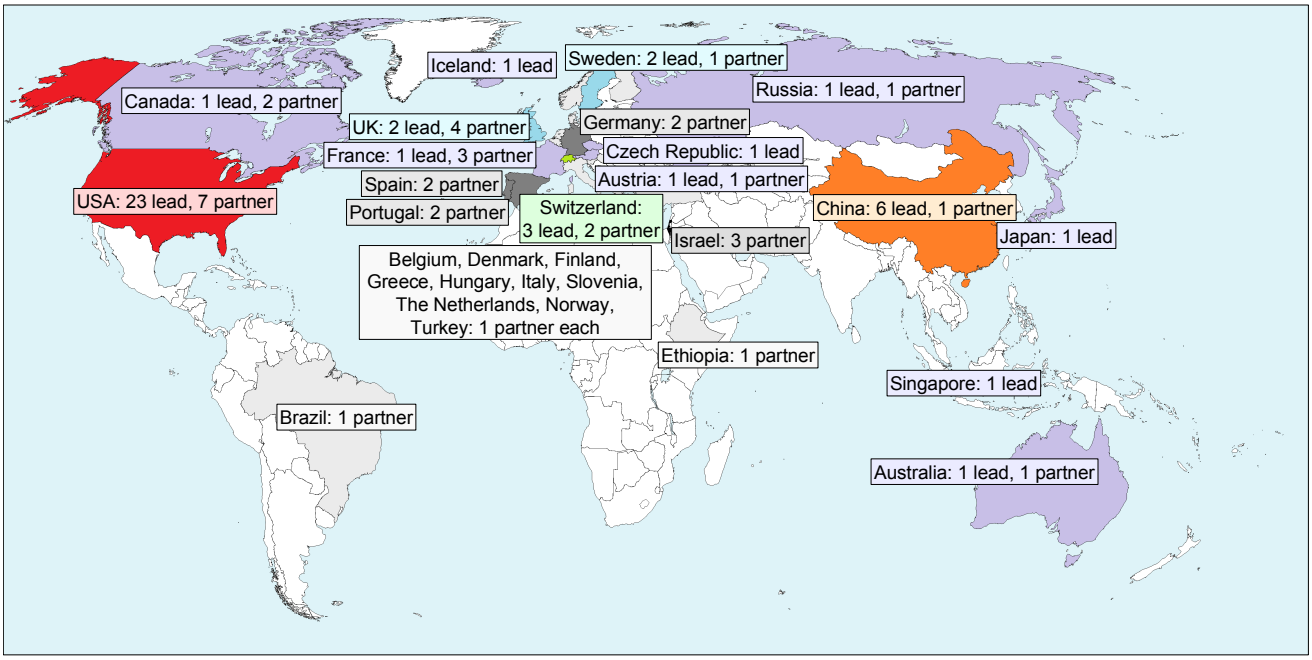

Figure 4. Map of nationality data. Any depictions of disputed territories are unintentional and do not indicate a position on the dispute.

The 23 US-based projects are based in 12 states and territories: 6 in California, 3 in New York, 2 in each of Massachusetts, Pennsylvania, and Tennessee, and 1 in each of Colorado, Michigan, Ohio, Oregon, Texas, Washington state, and the US Virgin Islands, with one project (CogPrime) not having a clear home state. This broad geographic distribution is due largely to the many academic projects: whereas US AI companies tend to concentrate in a few metropolitan areas, US universities are scattered widely across the country. Indeed, only two of the six California projects are academic. The for-profit projects are mainly in the metropolitan areas of Austin, New York City, Portland (Oregon), San Francisco, and Seattle, all of which are AI hotspots; the one exception is Victor in the US Virgin Islands. Additionally, OpenAI (a nonprofit) and Becca (no institutional home) are in the San Francisco and Boston areas, respectively, due to the AI industries in these cities.

13 projects are multinational. AERA, AIDEUS, Baidu Research, CommAI, Maluuba, Tencent AI Lab, and Uber AI Labs are each in two countries total (including the lead country). Animats, CogPrime, DeepMind, and SOAR are each in three countries. Blue Brain is in five countries. SingularityNET is in 8 countries. The Human Brain Project is in 19 countries, all of which are in Europe except for Israel and (depending on how one classifies it) Turkey. The most common international collaboration is UK-US, with both countries participating in four projects (Blue Brain, DeepMind, Soar, and Uber AI Labs). China and the US both participate in four (Baidu Research, CogPrime, SingularityNET, and Tencent AI Lab).

In geopolitical terms, there is a notable dominance of the US and its allies. Only five countries that are not US allies have AGI R&D projects: Brazil, China, Ethiopia, Russia, and (arguably) Switzerland, of which only China and Russia are considered US adversaries. Of the eight projects that China and Russia participate in, one is based in the US (CogPrime), three others have US participation (Baidu Research, SingularityNET, and Tencent AI Lab), and two are small projects with close ties to Western AI communities (AIDEUS, Real AI). The remaining two are projects of the Chinese Academy of Sciences that focus on basic neuroscience and medicine (China Brain Project, RCBII). The concentration of projects in the US and its allies, as well as the Western and academic orientation of the other projects, could make international governance of AGI R&D considerably easier.

In geopolitical terms, there is also a notable absence of some geopolitically important regions. There are no projects in South Asia, and just one (CogPrime) that is in part in Africa and one (SingularityNET) that is in part in Latin America. The AGI R&D in Russia consists mainly of the contributions of Alexey Potapov, who contributes to both AIDEUS and SingularityNET. Additionally, Potapov's published AGI research is mainly theoretical, not R&D (e.g., Potapov et al. 2016).[8] In Latin America, the only AGI R&D is the contributions of Cassio Pennachin to SingularityNET. Thus, the AGI R&D projects identified in this survey are almost all being conducted in institutions based in North America, Europe, the Middle East, East Asia, and Australia.

One important caveat is that no partner institutions for ACT-R were included. The ACT-R website lists numerous collaborating institutions, almost all of which are academic, spread across 21 countries. Several of these countries are not involved in other AGI R&D projects: India, Iran, Morocco, Pakistan, South Korea, Sri Lanka, and Venezuela. These partner institutions are excluded because this part of the ACT-R website is out of date. There may be some ongoing contributions to ACT-R in these countries; whether or not there are is beyond the scope of this paper.

Another important caveat comes from the many projects with open-source code. This code enables AGI R&D to be conducted anywhere in the world. It is thus possible that there are other countries involved in AGI R&D, perhaps a large number of other countries. The identification of countries whose participation consists exclusively of contributions to open-source code is beyond the scope of this paper.

5.6 Stated Goal

The dominant trend among stated goals is the preponderance of humanitarianism and intellectualism. 23 projects stated intellectualist goals and 20 stated humanitarian goals, while only 3 stated ecocentric goals (SingularityNET, Susaro, Victor), 2 stated profit goals (Maluuba, NNAISENSE), 2 stated animal welfare goals (Human Brain Project, proposing brain simulation to avoid animal testing, and SingularityNET, seeking to benefit all sentient beings), and 1 stated transhumanist goals (SingularityNET, seeking to benefit sentient beings and robots). 12 projects had unspecified goals.

[8] Also in Russia is Max Talanov of Kazan State University, who has contributed to NARS but not enough to be coded as a partner (e.g., Wang et al. 2016).

Some projects stated multiple goals: 11 projects stated 2 goals, 1 project (Human Brain Project) stated 3, and 1 project (SingularityNET) stated 4. Figure 5 summarizes the stated goal data.

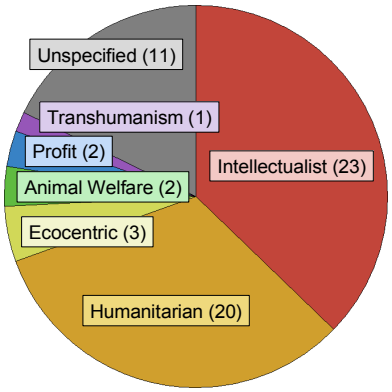

Figure 5. Summary of stated goal data. The figure shows more than 45 data points because some projects have multiple stated goals.

The intellectualist projects are predominantly academic, consistent with the wider emphasis on intellectual merit across academia. The intellectualist projects include 14 of the 19 projects based at least in part at an academic institution. (Each of the other five academic projects are coded as unspecified; they are generally smaller projects and are likely to have intellectualist goals.) The intellectualist projects also include the two projects at the Chinese Academy of Sciences. The Academy is coded as a government institution but is heavily academic in character. The other intellectualist projects are all in for-profit corporations, except AIDEUS, which has no institution type. Four of these corporate projects have close ties to academia (GoodAI, HTM, Maluuba, and NNAISENSE), as does AIDEUS. The fifth, MSR AI, is a bit of an outlier.

All eight US academic projects with military connections state intellectualist goals. Only one of them (LIDA) also states humanitarian goals. (The one non-US project with military connections, DSO-CA, had unspecified goals.) Only one other US academic project states humanitarian goals (NARS); the rest are unspecified. Meanwhile, three of eight non-US academic projects state humanitarian goals. One possible explanation for this is that the preponderance of military funding for US academic AGI R&D prompts projects to emphasize intellectual goals instead of humanitarian goals, whereas the availability of funding in other countries (especially across Europe) for other AGI applications, especially healthcare, prompts more emphasis on humanitarian goals. In short, the data may reflect the large US military budget and the large European civil budget, while also indicating that AGI researchers struggle to articulate military R&D in humanitarian terms.

10 of 18 for-profit projects state humanitarian goals. Many of these are Western (especially American) projects with a strong public face (e.g., CommAI/Facebook, DeepMind/Google, MSR AI/Microsoft, Nigel/Kimera, Uber AI Labs). Some of the other humanitarian for-profit projects are rooted in Western conceptions of altruism (e.g., GoodAI, Real AI). In contrast, the non-humanitarian for-profits are mostly smaller projects (e.g., Animats, NNAISENSE) and Chinese projects (Baidu Research, Tencent AI Lab). This suggests that the for-profit humanitarianism is mainly a mix of Western values and Western marketing.

The distinction between values and marketing is an important one and speaks to a limitation of this survey's methodology. Projects may publicly state goals that are appealing to their audiences while privately holding different goals. This may explain why ten for-profit projects state humanitarian goals while only two state profit goals. Some for-profit projects may genuinely care little about profit—indeed, two of them, GoodAI and Vicarious, explicitly reject profit as a goal. But others may only articulate humanitarianism to make themselves look good. This practice would be analogous to the

practice of "greenwashing", in which environmentally damaging corporations promote small activities that are environmentally beneficial to create the impression that they are more "green" than they actually are (e.g., Marquis et al. 2016). For example, a coal company might promote employees' in-office recycling to show their environmental commitment. Likewise, corporate AGI R&D projects may advertise humanitarian concern while mainly seeking profit, regardless of overall humanitarian consequences. One might refer to such conduct as "bluewashing", blue being the color of the United Nations and the World Humanitarian Summit.

Notable absences in the stated goals data include military advantage and the selfish benefit of AGI builders. Both of these is considered as an AGI R&D goal in prior literature, as is transhumanism/cosmism, which only gets brief support in one project (SingularityNET). The reason for these absences is beyond the scope of this paper, but some possibilities are plausible. Transhumanism and cosmism are not widely held goals across contemporary human society, though they are relatively common among AGI developers (e.g., Goertzel 2010). It is plausible that transhumanists and cosmists (outside of SingularityNET) prefer to keep their views more private so as to avoid raising concerns that their AGI R&D projects could threaten humanity. The pursuit of AGI for military advantage could raise similar concerns and could also prompt adversaries to commence or hasten their own AGI R&D. Finally, the pursuit of AGI for selfish benefit is antisocial and could pose reputational risks or prompt regulation if stated openly. Yet it is also possible that the people currently drawn to AGI R&D tend to actually have mainly humanitarian and intellectualist goals and not these other goals (Goertzel being a notable exception).

An important question is the extent to which AGI R&D projects share the same goals. Projects with different goals may be more likely to compete against each other to build AGI first (Armstrong et al. 2016). The preponderance of humanitarianism and intellectualism among AGI R&D projects suggests a considerable degree of consensus on goals. Furthermore, these goals are agent-neutral, further suggesting a low tendency to compete. But competition could occur anyway. One reason is that there can be important disagreements about the details of these particular views, such as in divergent conceptions of human rights between China and the West (Posner 2014). Additionally, even conscientious people can feel compelled to be the one to build AGI first, perhaps thinking to themselves "Forget about what's right and wrong. You have a tiger by the tail and will never have as much chance to influence events. Run with it!"[9] And of course, there can be disagreements between AGI humanitarians and intellectualists, as well as with the other goals that have been stated.

Finally, it should be noted that the coding of stated goals was especially interpretative. Many projects do not state their goals prominently or in philosophically neat terms. For example, DeepMind lists climate change as an important application. This could be either ecocentric or humanitarian or both, depending on why DeepMind seeks to address climate change. It was coded as humanitarian because it was mentioned in the context of "helping humanity tackle some of its greatest challenges", but it is plausible that ecocentrism was intended.

5.7 Engagement on Safety

Engagement on safety could only be identified for 17 projects. 12 of these projects were found to be active on safety (AERA, AIDEUS, CogPrime, DeepMind, FLOWERS, GoodAI, LIDA, NARS, OpenAI, Real AI, Susaro, and WBAI), 3 are moderate (CommAI, Maluuba, and Vicarious), and 2 are dismissive (HTM, Victor). Figure 6 summarizes the engagement on safety data.

[9] This quote is from cryptographer Martin Hellman, himself an evidently conscientious person. The quote refers to an episode in 1975 when Hellman debated whether to go public with information about breaking a code despite warnings from the US National Security Agency that doing so would be harmful. Hellman did go public with it, and in retrospect he concluded that it was the right thing to do, but that he had the wrong motivation for doing so (Hellman and Hellman 2016, p.48-50).

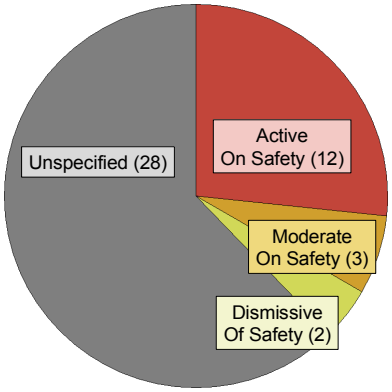

Figure 6. Summary of engagement on safety data.

One major trend in the engagement on safety data is the lack of attention to safety. Engagement on safety could not be identified for 28 of 45 projects. This provides some empirical support for the common assumption in prior AGI policy literature of AGI developers who want to proceed with inadequate regard for safety (Section 2.3). This survey's focus on publicly available data may overstate the neglect of safety because some projects may pay attention to safety without stating it publicly. For example, Animats and NNAISENSE were coded as unspecified, but there is little publicly available information about any aspect of these projects. Meanwhile, NNAISENSE Chief Scientist Jürgen Schmidhuber and NNAISENSE co-founder and Animats contributor Bas Steunebrink have done work on AGI safety (Steunebrink et al. 2016). Still, the data is strongly suggestive of widespread neglect of safety among AGI R&D projects.

Among the 17 projects for which engagement on safety was identified, some further trends are apparent. These 17 projects include 6 of the 8 projects with purely humanitarian goals and the only project with purely ecocentric goals yet only 1 of the 10 projects with purely intellectualist goals and 1 of 11 with unspecified goals. The 17 projects also include 9 of the 16 projects based purely at a for-profit and 3 of 4 projects based purely at a nonprofit yet only 2 of 15 projects based purely at an academic institution and 1 of 4 based in part at an academic institution. This suggests a cluster of projects that are broadly engaged on the impacts of AGI R&D, including ethics questions about what the impacts should be and risk/safety questions about whether the desired impacts will accrue, a cluster that is located predominantly outside of academia. Meanwhile, there is also a cluster of projects that are predominantly academic and view AGI in largely intellectual terms, to the extent that they state any goal at all. These trends suggest the importance of proposals to strengthen risk and ethics practices sometimes suggesting that adequate training could make AGI safety to be a reasonably tractable task among academic AGI R&D projects, such as via research review boards (Yampolskiy and Fox 2013).

Among the 12 projects found to be active on safety, a few themes are apparent. Several projects emphasized the importance of training AGI to be safe (AERA, CogPrime, GoodAI, and NARS), and critiquing concerns that AGI safety could be extremely difficult (Section 2.2; Goertzel and Pitt 2012; Goertzel 2015; 2016; Bieger al. 2015; Steunebrink et al. 2016).[10] Other projects focus on safety issues in near-term AI in the context of robotics (FLOWERS; Oudeyer et al. 2011) and reinforcement learning (DeepMind and OpenAI; Christiano et al. 2017), the latter being consistent with an agenda of using near-term AI to study long-term AI safety that was proposed by Google and OpenAI researchers (Amodei et al. 2016). LIDA has explored fundamentals of AGI morality as they relate to engineered systems (Wallach et al. 2010; 2011; Madl and Franklin 2015). WBAI suggests that it seeks to build brain-like AGI in part because its similarity to human intelligence would make it safer.

[10] Similarly, the "safety-moderate" project Vicarious states a belief that AGI safety may not be difficult, prompting them to be not (yet) active on safety.

5.8 Size

Projects were found to have size mainly in the small to medium range. 13 projects were coded as small, 11 as medium-small, 12 as medium, 5 as medium-large (Blue Brain, CogPrime, MSR AI, Soar, and Vicarious), and 3 as large (DeepMind, Human Brain Project, and OpenAI), with 1 project unspecified (Susaro, a project in "stealth mode" with no size information found). Figure 7 summarizes the size data.

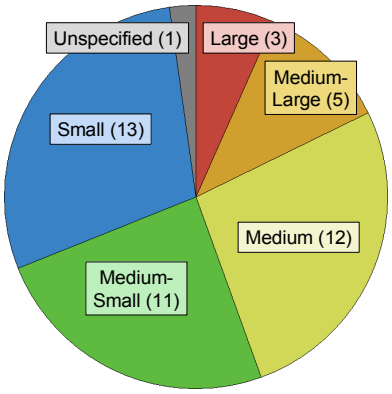

Figure 7. Summary of size data.

While the projects at each size point are diverse, some trends are nonetheless apparent. With respect to institution type, academic projects tend to be somewhat smaller, while corporate projects tend to be somewhat larger, though there are both large academic projects and small corporate projects. 7 of the 13 small projects are academic, while only three are in private corporations and none are in public corporations. 3 of the 8 medium-large or large projects are academic, while 4 are in corporations.

The size vs. institution type trends may be in part a coding artifact, because academic and corporate projects have different indicators of size. Corporate projects are less likely to list their full personnel or to catalog their productivity in open publication lists. Corporate projects may also have substantial portions focused on near-term applications at the expense of long-term AGI R&D, though academic projects may similarly have portions focused on intellectual goals at the expense of AGI R&D. Corporate projects, especially those in public corporations, often have additional resources that can be distributed to AGI projects, including large datasets and funding. Academic institutions can find funding for AGI projects, but generally not as much as corporations, especially the large public corporations. This distinction between academic and corporate projects is illustrated, for example, by NNAISENSE, which was launched by academics in part to prevent other companies from poaching research talent. It further suggests that the largest AGI R&D projects may increasingly be corporate, especially if there is AGI R&D-profit synergy (Section 5.2).

With respect to open-source, there is a clear trend towards larger projects being more open-source. Of the 17 projects that do not make source code available, 8 are small and 6 are medium-small, while only 1 is medium and 1 is medium-large, with 1 being of unspecified size. In contrast, the 24 unrestricted open-source projects include 5 of 11 small projects, 5 of 11 medium-small projects, 9 of 12 medium projects, 4 of 5 medium-large projects, and 2 of 3 large projects.

With respect to military connections, the projects with military connections tend to be in the small to medium range. The 8 military projects include 2 small, 2 medium-small, 4 medium, and 1 medium-large. This makes sense given that these are primarily US academic projects with military funding. Academic projects tend to be smaller, while those with external funding tend to be more medium in

size. Academic projects that forgo military funding may sometimes be smaller than they otherwise could have been.

With respect to nationality, there is some trend towards US-based projects being larger while China-based projects are smaller, though, for the US, it is a somewhat weak trend. The 23 US-based projects include 6 of 13 small projects, 3 of 11 medium-small projects, 9 of 12 medium projects, 4 of 5 medium-large projects, and 1 of 3 large projects. In contrast, all six China-based projects are either small or medium-small. This trend strengthens the Section 5.5 finding that AGI is concentrated in the US and its allies. An important caveat is that two of the Chinese projects are based in large Chinese corporations, Baidu and Tencent. These corporations have large AI groups that only show a small amount of attention to AGI. If the corporations decide to do more in AGI, they could scale up quickly. It is also possible that they are already doing more, in work not identified by this survey, though the same could also be said for projects based in other countries.

With respect to stated goal, a few trends can be discerned. First, the projects with unspecified goals tend to be small, including 7 of 13 small projects. This makes sense: smaller projects have less opportunity to state their goals. Intellectualist projects tend to be medium-small to medium, including 6 of 11 medium-small projects and 9 of 12 medium projects, which is consistent with the intellectualist projects tending to be academic. Humanitarian projects tend to be larger, including 6 of 12 medium projects, 3 of 5 medium-large projects, and all 3 large projects. A possible explanation is that the larger projects tend to have wider societal support, prompting the projects to take a more humanitarian position. The preponderance of humanitarianism among the larger projects could facilitate the development of consensus on goals among the projects that are most likely to build AGI. Such consensus could in turn help to avoid a risky race dynamic (Armstrong et al. 2016).

Finally, with respect to engagement on safety, there is a weak trend towards the larger projects being more active on safety. The active projects include 2 small, 3 medium-small, 3 medium, 1 medium-large, 2 large, and 1 of unspecified size. In contrast, the projects with unspecified engagement on safety include 10 small, 7 medium-small, 6 medium, 3 medium-large, and 1 large. Thus, projects of all sizes can be found active or not active on safety, though the larger projects do have a somewhat greater tendency to be active.

5.9 Other Notable Projects

47 other notable projects were recorded in the process of identifying AGI R&D projects. These include 26 inactive AGI R&D projects, 9 projects that are not AGI, and 12 projects that are not R&D. Unlike with active AGI R&D projects, no attempt was made to be comprehensive in the identification of other notable projects. It is likely that some notable projects are not included. The list of other notable projects and brief details about them are presented in Appendix 2.

The 26 inactive projects are mainly academic, such as 4CAPS (led by Marcel Just of Carnegie Mellon University), Artificial Brain Laboratory (led by Hugo de Garis of Xiamen University), CERA-CRANIUM (led by Raúl Arrabales of University of Madrid), CHREST (led by Fernand Gobet of University of Liverpool), DUAL (led by Boicho Kokinov of New Bulgarian University), EPIC (led by David Kieras at University of Michigan), OSCAR (led by John Pollock of University of Arizona), and Shruti (led by Lokendra Shastri of University of California, Berkeley). They varied considerably in duration, from a few years (e.g., AGINAO, active from 2011 to 2013) to over a decade (e.g., CHREST, active from 1992 to 2012).

The nine projects that are not AGI are mostly AI groups at large computing technology corporations. Six corporations were searched carefully for AGI projects and found not to have any:

Alibaba, Amazon, Apple, Intel, SalesForce, and Twitter.[11] Given these corporations' extensive resources, it is notable that they do not appear to have any AGI projects.[12] In addition to DeepMind, two other projects at Google were considered: Google Brain and Quantum AI Lab. While Google Brain has done some AGI work with DeepMind, it focuses on narrow AI. Finally, one narrow cognitive architecture project was included (Xapagy) as an illustrative example. Many more narrow cognitive architecture projects could have been included—Kotseruba et al. (2016) lists 86 cognitive architecture projects, most of which are narrow.

The 12 projects that are not R&D cover a mix of different aspects of AGI. Some focus on basic science, including several brain projects (e.g., the BRAIN Initiative at the US National Institutes of Health and Brain/MINDS at Japan's Ministry of Education, Culture, Sports, Science, and Technology). Several focus on hardware and software for building AGI (e.g., the IBM Cognitive Computing Project, the Cognitive Systems Toolkit project led by Ricardo Gudwin of University of Campinas in Brazil, and the Neurogrid project led by Kwabena Boahen of Stanford University). Two focus on safety aspects of AGI design (Center for Human-Compatible AI at University of California, Berkeley and the Machine Intelligence Research Institute). One (Carboncopies) focuses on supporting other R&D projects. Finally, one focuses on theoretical aspects of AGI (Goedel Machine, led by Jürgen Schmidhuber of the Dalle Molle Institute for Artificial Intelligence Research in Switzerland). This is not a comprehensive list of projects working on non-R&D aspects of AGI. For example, projects working on AGI ethics, risk, and policy were not included because they are further removed from R&D.

6. Conclusion

Despite the seemingly speculative nature of AGI, R&D towards building it is already happening. This survey identifies 45 AGI R&D projects spread across 30 countries in 6 continents, many of which are based in major corporations and academic institutions, and some of which are large and heavily funded. Given that this survey relies exclusively on openly published information, this should be treated as a lower bound for the total extent of AGI R&D. Thus, regardless of how speculative AGI itself may be, R&D towards it is clearly very real. Given the potentially high stakes of AGI in terms of ethics, risk, and policy, the AGI R&D projects warrant ongoing attention.

6.1 Main Findings

Regarding ethics, the major trend is projects' split between humanitarian and intellectualist goals, with the former coming largely from corporate projects and the latter from academic projects. There is reason to be suspicious of corporate statements of humanitarianism—they may be "bluewashing" (Section 5.6) to conceal self-interested profit goals. Still, among the projects not motivated by intellectual goals, there does seem to be a bit of a consensus for at least some form of humanitarianism, and not for other types of goals commonly found in AGI discourses, such as cosmism/transhumanism. Meanwhile, the intellectualist projects indicate that academic R&D projects still tend to view their work in intellectual terms, instead of in terms of societal impacts or other ethical factors, even for potentially high-impact pursuits like AGI.

Regarding risk, two points stand out. First, a clear majority of projects had no identifiable engagement on safety. While many of these projects are small, it includes even some of the larger projects. It appears that concerns about AGI risk have not caught on across much of AGI R&D,

[11] IBM was also carefully searched. No AGI R&D was found at IBM, but IBM does have a project doing hardware development related to AGI, the Cognitive Computing Project, which is included in Appendix 2.

[12] It is possible that some of them have AGI projects that are not publicly acknowledged. Apple in particular has a reputation for secrecy, making it a relatively likely host of an unacknowledged AGI project.

especially within academia. Second, some trends suggest that a risky race dynamic may be avoidable. One is the concentration of projects, especially larger projects, in the US and its allies, which can facilitate both informal coordination and formal public policy. Another is the modest consensus for humanitarianism, again especially among larger projects, which could reduce projects' desire to compete against each other. Finally, many of the projects are interconnected via shared personnel, parent organizations, AGI systems development, and participating in the same communities, such as the AGI Society. This suggests a community working together towards a common goal, not competing against each other to "win".

Regarding policy, several important points can be made. One is the concentration of projects in the US and its allies, including most of the larger projects (or all of them, depending on which countries are considered US allies). This could greatly facilitate the establishment of AGI policy with jurisdiction for most AGI R&D projects, including all of the larger ones. Another important point is the concentration of projects in academia and corporations, with relatively little in government or with military connections. Each institution type merits its own type of policy, such as review boards in academia and financial incentive structures for corporations. The potential for AGI R&D-profit synergy (Section 5.2) could be especially important here, determining both the extent of corporate R&D and the willingness of corporations to submit to restrictive regulations. This survey finds hints of AGI R&D-profit synergy, but not the massive synergies found for certain other types of AI. Finally, the preponderance of projects with at least some open-source code complicates any policy effort, because the R&D could in principle be done by anyone, anywhere.

6.2 Limitations and Future Work

The above conclusions seem robust given this study's methodology, but other methodologies could point in different directions. For example, the published statements about goals suggest a consensus towards humanitarian and intellectualist goals, but this could miss unpublished disagreements on the specifics of these goals. One potential flashpoint is if Western humanitarian AGI projects seek to advance political freedom, whereas Chinese AGI projects seek to advance economic development, in parallel with the broader disagreement about human rights between China and the West. Furthermore, the published statements about goals used in this survey could deviate from projects' actual goals if they are not entirely honest in their published statements. Projects may be proceeding recklessly towards selfish goals while presenting a responsible, ethical front to the public. These possibilities suggest a more difficult policy challenge and larger AGI risk. Alternatively, this study could overestimate the risk. Perhaps many projects have similar goals and concerns about safety, even if they have not published any statements to this effect. Future research using other methodologies, especially interviews with people involved in AGI R&D, may yield further insights.

A different complication comes from the possibility that there are other AGI R&D projects not identified by this study. Some projects may have a public presence but were simply not identified in this study's searches, despite the efforts made to be comprehensive. This study is especially likely to miss projects that work in non-English languages, since it only conducted searches in English. However, English is the primary international language for AI and for science in general, so it is plausible that no non-English projects were missed.

Furthermore, there may be additional AGI R&D projects that have no public face at all. This could conceivably include the secret government military projects that some might fear. However, the modest nature of the military connections of projects identified in this survey suggests that there may be no major military AGI projects at this time. Specifically, the projects identified with military connections are generally small and focused on mundane (by military standards) tactical issues, not grand ambitions of global conquest. This makes it likely that there are not any more ambitious secret

military AGI projects at this time. However, given the stakes, it is important to remain vigilant about the possibility.

Another, more likely possibility is of stealth mode private sector projects. This study identified one stealth mode project that happens to have a public website (Susaro); perhaps there are others without websites. Some of these may be independent startups, while others may be projects within larger AI organizations. The larger organizations in particular often have resources to support AGI R&D and could camouflage it within other AI projects. Some larger AI organizations, such as Apple, have reputations for secrecy and may be likely candidates for hosting stealth AGI projects. The possibility of stealth projects or other projects not identified in this survey means that the number of projects identified in this survey (45) should be treated as a lower bound.

A different potential source of additional AGI R&D projects is the vast space of projects focused on deep learning and related techniques. These projects were excluded from this survey because there are too many to assess in this study's project-by-project methodology and because there are diverging opinions on whether these projects rate as AGI R&D. If AGI could come from deep learning and related techniques, the AGI R&D landscape would look substantially different from the picture painted in this paper, with major consequences for ethics, risk, and policy. Therefore, an important direction for future research is to assess the possibility that AGI could come from deep learning and related techniques and then relate this to ethics, risk, and policy.

Another worthwhile direction for future research is on projects' capability to build AGI. This study includes project size as a proxy for capability, but it is an imperfect proxy. Capability is the more important attribute for understanding which projects may be closest to building AGI. More capable projects may warrant greater policy attention, and a cluster of projects with similar capability could lead to a risky race dynamic. (Project size is important for other reasons, such as projects' pull on the labor market for AGI researchers or their potential for political lobbying.) Project capacity could be assessed via attention to details of projects' performance to date, the viability of their technical approaches to AGI, and other factors. Given the ethics, risk, and policy importance of project capacity, this is an important direction for future research.

Finally, future research could also investigate other actors involved in AGI. In addition to the R&D projects, there are also, among others, R&D groups for related aspects of AGI, such as hardware and safety measures; people studying and working on AGI ethics, risk, and policy; and "epistemic activists" promoting certain understandings of AGI. Each of these populations can play significant roles in ultimate AGI outcomes and likewise has implications for ethics, risk, and policy that can be worth considering. Empirical study of these populations could clarify the nature of the work being done, identify gaps, and suggests trends in how AGI could play out.

6.3 Concluding Remarks

Overall, the present study shows that AGI ethics, risk, and policy can be pursued with a sound empirical basis—it need not be based solely on speculation about hypothetical AGI R&D projects and actors. The present study additionally makes progress on this empirical basis by contributing the first-ever survey of AGI R&D projects for ethics, risk, and policy. Given the potentially high stakes of AGI, it is hoped that this research can be used productively towards improving AGI outcomes.

Appendix 1. Active AGI R&D Projects

ACT-R

Main website: http://act-r.psy.cmu.edu

ACT-R is a research project led by John Anderson of Carnegie Mellon University. It is a theory of human cognition and a computational framework for simulating human cognition.[13] ACT-R is an acronym for Adaptive Control of Thought-Rational. It was briefly connected to Leabra via the SAL project.

Lead institutions: Carnegie Mellon University

Partner institutions: none

- The ACT-R website lists numerous collaborating institutions across 21 countries,[14] though this includes people who previously contributed and have since moved on to other projects, and does not include some co-authors of recent papers.[15] No active partner institutions could be confirmed from the website and thus none are coded here, though there may be active partner institutions.

Type of institution: academic

Open-source: yes[16]

Military connection: yes[17]

Lead country: USA

Partner countries: none

Stated goals: intellectualism

- The main description of ACT-R on its website is exclusively about the intellectual research, with no mention of broader impacts.[18]

Engagement on safety: unspecified

Size: medium

[13] http://act-r.psy.cmu.edu/about
[14] http://act-r.psy.cmu.edu/people
[15] For example, Lee et al. (2017) has lead author Hee Seung Lee of Yonsei University, who is not listed at http://act-r.psy.cmu.edu/people.
[16] http://act-r.psy.cmu.edu/software
[17] US Office of Naval Research funding is reported in Zhang et al. (2016).
[18] http://act-r.psy.cmu.edu/about

AERA

Main website: http://www.ru.is/faculty/thorisson

AERA is led by Kristinn Thórisson of Reykjavik University. AERA is an acronym for Autocatalytic Endogenous Reflective Architecture (Nivel et al. 2013). The project aims "to both understand the mind and build a practical AGI system", and is currently being used "to study machine understanding, teaching methodologies for artificial learners, even the development of ethical values".[19]

Lead institutions: Reykjavik University

Partner institutions: Icelandic Institute for Intelligent Machines (Iceland), Dalle Molle Institute for Artificial Intelligence Research (Switzerland) (per authors listed in Steunebrink et al. 2016)

Type of institution: academic

Open-source: no

Military connection: no

- Project lead Thórisson criticizes military AI in Icelandic Institute for Intelligent Machines (IIIM).[20]

Lead country: Iceland

Partner countries: Switzerland

Stated goals: humanitarianism, intellectualism

- Thórisson's website links to the ethics policy of IIIM, which aims to "to advance scientific understanding of the world, and to enable the application of this knowledge for the benefit and betterment of humankind", with emphasis on concerns about privacy and military misuse.[21]
- The ethics policy also states an aim "to focus its research towards topics and challenges of obvious benefit to the general public, and for the betterment of society, human livelihood and life on Earth". This mention of life on Earth suggests ecocentrism, but all other text is humanitarian or intellectualist.[22]

Engagement on safety: active

- The AERA group has written on how to enhance safety during AGI self-improvement, arguing that certain design principles would make it easy to achieve safety (Steunebrink et al. 2016).
- Thórisson's website links to an article by AI researcher Oren Etzioni (2016) that is dismissive of concerns about AGI, suggesting that AERA may also be dismissive, but this could not be concluded from just the presence of the link.

Size: medium-small

[19] http://www.ru.is/faculty/thorisson
[20] http://www.iiim.is/ethics-policy/
[21] http://www.iiim.is/ethics-policy/3
[22] http://www.iiim.is/ethics-policy/3

AIDEUS

Main website: http://aideus.com

AIDEUS is led by Alexey Potapov of ITMO University in Saint Petersburg and Sergey Rodionov of Aix Marseille Université. Potapov is also an advisor to SingularityNET.[23] It states a goal of the "creation of a strong artificial intelligence".[24] Its approach is to "proceed from universal prediction models on the basis of algorithmic probability used for choosing optimal actions".[25] It has published frequently on AGI, often in the proceedings of AGI conferences.[26] Recent publications report funding from the Russian Federation, including the Ministry of Education and Science (e.g., Potapov et al. 2016).

Lead institutions: AIDEUS

Partner institutions: none

Type of institution: none

- No institution type is specified on the project website. AIDEUS is listed separately from academic affiliations on publications (e.g., Potapov et al. 2016). It is listed as a company on its Facebook page,[27] but the Facebook "company" category is not restricted to corporations.

Open-source: no

Military connection: unspecified

- Funding sources include "Government of Russian Federation, Grant 074-U01", which does not appear to be military, but this could not be confirmed.

Lead country: Russia

Partner countries: France

Stated goals: humanitarian, intellectualist

- The project aims to build superintelligence in order to "help us better understand our own thinking and to solve difficult scientific, technical, social and economic problems."[28]

Engagement on safety: active

- AIDEUS has published AGI safety research, e.g. Potapov and Rodionov (2014).

Size: small

[23] https://singularitynet.io
[24] http://aideus.com
[25] http://aideus.com/research/research.html
[26] http://aideus.com/research/publications.html
[27] https://www.facebook.com/pg/Aideus-Strong-artificial-intelligence-455322977847194/about
[28] http://aideus.com/community/community.html

AIXI

Main website: http://www.hutter1.net/ai/aixigentle.htm

AIXI is led by Marcus Hutter of Australian National University. AIXI is based on a "meta-algorithm" that searches the space of algorithms to find the best one for AGI.[29] Hutter proves that AIXI will find the most intelligent AI if given infinite computing power. While this is purely a theoretical result, it has led to approximate versions that are implemented in computer code.[30]

Lead institutions: Australian National University

Partner institutions: none

Type of institution: Academic

Open-source: no

Military connection: unspecified

Lead country: Australia

Partner countries: none

Stated goals: unspecified

Engagement on safety: unspecified

Size: small

[29] Goertzel (2014, p.25)
[30] Veness et al. (2011)

Alice In Wonderland (AIW)

Main website: https://github.com/arnizamani/aiw and http://flov.gu.se/english/about/staff?languageId=100001&userId=xstrcl

AIW is led by Claes Strannegård of Chalmers University of Technology in Sweden. A paper about AIW in *Journal of Artificial General Intelligence* describes it as being similar to NARS (Strannegård et al. 2017a). A separate paper describes it as a prototype for implementing new ideas about "bridging the gap between symbolic and sub-symbolic reasoning" (Strannegård et al. 2017b).

Lead institutions: Chalmers University of Technology

Partner institutions: none

Type of institution: academic

Open-source: yes[31]

Military connection: no[32]

Lead country: Sweden

Partner countries: none

Stated goals: unspecified

Engagement on safety: unspecified

Size: small

[31] https://github.com/arnizamani/aiw
[32] Reported funding is from the Swedish Research Council

Animats

Main website: https://github.com/ni1s/animats

Animats is a small project developed for the First International Workshop on Architectures for Generality & Autonomy[33] and the 2017 AGI conference.[34] The project is a collaboration between researchers at universities in Sweden and the United States and the Swiss company NNAISENSE.[35] It seeks to build AGI based on animal intelligence.[36]

Lead institutions: Chalmers University of Technology

Partner institutions: University of Gothenburg, Harvard University, NNAISENSE

Type of institution: academic, private corporation

Open-source: yes[37]

Military connection: no[38]

Lead country: Sweden

Partner countries: Switzerland, USA

Stated goals: unspecified

Engagement on safety: unspecified

Size: small

[33] http://cadia.ru.is/workshops/aga2017
[34] Strannegård et al. (2017)
[35] Strannegård et al. (2017b)
[36] Strannegård et al. (2017b)
[37] https://github.com/ni1s/animats
[38] Reported funding is from the Swedish Research Council

Baidu Research

Main website: http://research.baidu.com/learning-speak-via-interaction

Baidu Research is an AI research group within Baidu. It has offices in Beijing, Shenzhen, and Sunnyvale, California.[39] One page of its website states that its mission is "to create general intelligence for machines",[40] though this does not appear to be a major theme for the group. It has achieved success in "zero-shot learning" in language processing, in which the AI "is able to understand unseen sentences".[41] Some observers rate this as a significant breakthrough.[42]

Lead institutions: Baidu

Partner institutions: none

Type of institution: Public Corporation

Open-source: no
- Baidu releases some work open-source,[43] but not its AGI

Military connection: unspecified
- Baidu receives AI funding from the Chinese government for "computer vision, biometric identification, intellectual property rights, and human-computer interaction".[44]

Lead country: China

Partner countries: USA

Stated goals: unspecified

Engagement on safety: unspecified
- Former Baidu Research Chief Scientist Andrew Ng says that Baidu takes safety and ethics seriously, and expresses his personal views that AI will help humanity and that "fears about AI and killer robots are overblown",[45] but this was not in the context of AGI. No direct discussion of safety by Baidu Research was found.

Size: medium-small

[39] http://bdl.baidu.com/contact_b.html
[40] http://research.baidu.com/learning-speak-via-interaction
[41] http://research.baidu.com/ai-agent-human-like-language-acquisition-virtual-environment
[42] Griffin (2017)
[43] https://github.com/baidu
[44] Gershgorn (2017)
[45] Maddox (2016)

Becca

Main website: https://github.com/brohrer/becca

Becca is led by Brandon Rohrer, currently at Facebook.[46] According to its website, Becca "is a general learning program for use in any robot or embodied system"; it "aspires to be a brain for any robot, doing anything".[47] Rohrer describes it as an open-source project,[48] and its source code is available on its website.

Lead institutions: Becca

Partner institutions: none

Type of institution: none
- No institutional home for Becca was found; it appears to be a spare-time project for Rohrer

Open-source: yes[49]

Military connection: unspecified
- Becca began while Rohrer was at Sandia National Laboratories,[50] but this connection appears to be inactive.

Lead country: USA

Partner countries: none

Stated goals: unspecified

Engagement on safety: unspecified

Size: small

[46] https://www.linkedin.com/in/brohrer
[47] https://github.com/brohrer/becca
[48] https://www.linkedin.com/in/brohrer
[49] https://github.com/brohrer/becca
[50] https://www.linkedin.com/in/brohrer

Blue Brain

Main website: http://bluebrain.epfl.ch

Blue Brain is a research project led by Henry Markram. It has been active since 2005. Its website states that its goal "is to build biologically detailed digital reconstructions and simulations of the rodent, and ultimately the human brain".[51] Markram also founded the Human Brain Project, which shares research strategy with Blue Brain.[52]

Lead institution: École Polytechnique Fédérale de Lausanne

Partner institutions: Hebrew University (Israel); Universidad Politecnica de Madrid and Consejo Superior de Investigaciones Científicas (Spanish National Research Council) (Spain); University of London (UK); and IBM, St. Elizabeth's Medical Center, University of Nevada-Reno, and Yale University (USA)[53]

Type of institution: academic

- Blue Brain lists the public corporation IBM as a contributor, making "its researchers available to help install the BlueGene supercomputer and set up circuits that would be adequate for simulating a neuronal network".[54] This contribution was judged to be too small to code Blue Brain as part public corporation.

Open-source: yes[55]

Military connection: unspecified

Lead country: Switzerland

Partner countries: Israel, Spain, UK, USA

Stated goals: humanitarianism, intellectualism

- The Blue Brain website states that "Understanding the brain is vital, not just to understand the biological mechanisms which give us our thoughts and emotions and which make us human, but for practical reasons," the latter including computing, robotics, and medicine.[56] These applications are broadly humanitarian, though it is a more muted humanitarian than what is found for several other projects.

Engagement on safety: unspecified

Size: medium-large

[51] http://bluebrain.epfl.ch/page-56882-en.html
[52] http://bluebrain.epfl.ch/page-52741-en.html
[53] http://bluebrain.epfl.ch/page-56897-en.html
[54] http://bluebrain.epfl.ch/page-56897-en.html
[55] https://github.com/BlueBrain
[56] http://bluebrain.epfl.ch/page-125344-en.html

China Brain Project

Main website: none found

China Brain Project is a research project of the Chinese Academy of Sciences focused on basic and clinical neuroscience and brain-inspired computing. As of July 2016, the Chinese Academy of Sciences had announced the project and said it would launch soon,[57] but in August 2017 no project website was found. Project lead Mu-Ming Poo described the project in a 2016 article in the journal *Neuron*, stating that "Learning from information processing mechanisms of the brain is clearly a promising way forward in building stronger and more general machine intelligence" and "The China Brain Project will focus its efforts on developing cognitive robotics as a platform for integrating brain-inspired computational models and devices".[58]

Lead institutions: Chinese Academy of Sciences

Partner institutions: none

Type of institution: government
- The Chinese Academy of Sciences is a public institution under the Chinese government.[59]
- Mu-Ming Poo lists the Chinese Natural Science Foundation and Ministry of Science and Technology as guiding organizations for the China Brain Project.[60]

Open-source: no

Military connection: unspecified

Lead country: China

Partner countries: none

Stated goals: humanitarian, intellectual
- Mu-Ming Poo describes the project's goals as "understanding the neural basis of human cognition" and "reducing the societal burden of major brain disorders"[61]

Engagement on safety: unspecified

Size: small

[57] http://english.cas.cn/newsroom/news/201606/t20160617_164529.shtml
[58] Poo et al. (2016)
[59] http://english.cas.cn/about_us/introduction/201501/t20150114_135284.shtml
[60] Poo et al. (2016)
[61] Poo et al. (2016)

CLARION

Main website: http://www.clarioncognitivearchitecture.com, http://www.cogsci.rpi.edu/~rsun/clarion.html

CLARION is led by Ron Sun of Rensselaer Polytechnic Institute in research dating to around 1994.[62] It is a hybrid of implicit and explicit knowledge, which Goertzel (2014, p.23) describes as a powerful approach but incomplete and not well integrated.

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Lead institutions: Rensselaer Polytechnic Institute

Partner institutions: none

Type of institution: academic

Open-source: yes[63]

Military connection: yes[64]

Lead country: USA

Partner countries: none

Stated goals: intellectualist

- The CLARION website states that it "aims to investigate the fundamental structures of the human mind", with "the ultimate goal of providing unified explanations for a wide range of cognitive phenomenon".[65]

Engagement on safety: unspecified

Size: medium-small

[62] http://www.cogsci.rpi.edu/~rsun/hybrid-resource.html
[63] http://www.cogsci.rpi.edu/~rsun/clarion.html
[64] US Office of Naval Research and Army Research Institute funding is reported at http://www.cogsci.rpi.edu/~rsun/clarion.html
[65] http://www.clarioncognitivearchitecture.com/home

CogPrime

Main website: http://wiki.opencog.org/w/CogPrime_Overview

CogPrime is an open-source project led by Ben Goertzel. Goertzel also leads SingularityNET. It "integrates multiple learning algorithms associated with different memory types" currently in use for controlling video game characters and planned for controlling humanoid robots.[66] The CogPrime website describes it is an approach "for achieving general intelligence that is qualitatively human-level and in many respects human-like".[67]

Lead institution: OpenCog Foundation

Partner institutions: none

Type of institution: nonprofit
- OpenCog is described as a nonprofit with 501(c)(3) status pending[68]

Open-source: yes[69]

Military connection: unspecified
- OpenCog has expressed concerns about military AGI.[70]

Lead country: USA
- While OpenCog is highly international, its website refers to an application for 501(c)(3) nonprofit status,[71] implying an administrative home in the US.

Partner countries: China, Ethiopia[72]
- Goertzel (2012b) writes that "OpenCog developers are working on multiple continents".

Stated goals: unspecified
- The CogPrime website describes CogPrime as able to implement a range of ethical views.[73] Elsewhere, CogPrime leader Goertzel has advocated cosmism (Goertzel 2010).

Engagement on safety: active
- CogPrime states that AGI "should be able to reliably achieve a much higher level of commonsensically ethical behavior than any human being", adding that "Our explorations in the detailed design of CogPrime's goal system have done nothing to degrade this belief".[74]

Size: medium-large

[66] Goertzel (2014, p.23)
[67] http://wiki.opencog.org/w/CogPrime_Overview
[68] http://opencog.org/about
[69] https://github.com/opencog
[70] http://blog.opencog.org/2014/01/17/what-is-consciousness
[71] http://opencog.org/about
[72] Goertzel (2008); http://opencog.org/2013/04/new-opencog-ai-lab-opens-in-addis-ababa-ethiopia
[73] http://wiki.opencog.org/w/CogPrime_Overview
[74] http://wiki.opencog.org/w/CogPrime_Overview

CommAI

Main website: https://research.fb.com/projects/commai

CommAI is a project of FAIR (Facebook AI Research). FAIR is led by Yann LeCun, who is also on the faculty at New York University. The FAIR website states that it seeks "to understand and develop systems with human level intelligence by advancing the longer-term academic problems surrounding AI".[75] CommAI is the FAIR project most focused on AGI. The CommAI group has published an initial paper "CommAI: Evaluating the First Steps Towards a Useful General AI".[76] CommAI is used in the General AI Challenge sponsored by GoodAI.[77]

Lead institution: Facebook

Partner institutions: none

Type of institution: public corporation

Open-source: yes[78]

Military connection: unspecified
- Facebook does not appear to have any US defense contracts.[79]

Lead country: USA

Partner countries: France
- FAIR based in New York City and has also has offices in Menlo Park, California and Paris.[80]

Stated goals: humanitarianism
- The CommAI website states that it is "aiming at developing general-purpose artificial agents that are *useful* for humans in their daily endeavours" (emphasis original).[81] Facebook is also a founding partner of the Partnership on AI to Benefit People & Society, which has humanitarian goals.[82]

Engagement on safety: moderate
- Facebook is a founding partner of the Partnership on AI to Benefit People & Society, which expresses concern about AI safety,[83] but CommAI shows no direct involvement on safety.

Size: medium

[75] https://research.fb.com/category/facebook-ai-research-fair
[76] https://research.fb.com/publications/commai-evaluating-the-first-steps-towards-a-useful-general-ai
[77] https://research.fb.com/projects/commai
[78] https://github.com/facebookresearch/CommAI-env
[79] https://www.fpds.gov/ezsearch/fpdsportal?q=facebook+DEPARTMENT_FULL_NAME%3A"DEPT+OF+DEFENSE"
[80] https://code.facebook.com/posts/447727685394738/facebook-expands-ai-research-team-to-paris
[81] https://research.fb.com/projects/commai
[82] https://www.partnershiponai.org/tenets
[83] https://www.partnershiponai.org/tenets

Cyc

Main website: http://www.cyc.com

Cyc is led by Doug Lenat, who began Cyc in 1984. Cyc is a project of Cycorp, a corporation based in Austin, Texas that uses Cyc for consulting and other services to other corporations and government agencies.[84] The Cycorp website describes Cyc as "a long-term quest to develop a true artificial intelligence".[85] Cyc has a unique database of millions of hand-coded items of commonsense human knowledge, which it aims to leverage for human-level AGI.[86] In an interview, Lenat says "Cycorp's goal is to *codify general human knowledge and common sense* so that computers might make use of it" (emphasis original).[87]

Lead institution: Cycorp

Partner institutions: none

Type of institution: private corporation

Open-source: restricted
- Cycorp offers a no-cost license for researchers upon request[88]
- Part of Cyc was briefly made available as OpenCyc, but this was discontinued[89]
- Cycorp "has placed the core Cyc ontology into the public domain".[90]

Military connection: yes
- Cycorp received a $25 million contract to analyze terrorism for the US military.[91]

Lead country: USA

Partner countries: none

Stated goals: unspecified

Engagement on safety: unspecified

Size: medium

[84] http://www.cyc.com/enterprise-solutions
[85] http://www.cyc.com/about/company-profile
[86] Goertzel 2014, p.16
[87] Love (2014)
[88] http://www.cyc.com/platform/researchcyc
[89] http://www.cyc.com/platform
[90] http://www.cyc.com/about/company-profile
[91] http://www.cyc.com/about/media-coverage/computer-save-world; see also Deaton et al. (2005)

DeepMind

Main website: http://deepmind.com

DeepMind is an AI corporation led by Demis Hassabis, Shane Legg, and Mustafa Suleyman. It was founded in 2010 and acquired by Google in 2014 for £400m ($650m; Gibbs 2014). It seeks to develop "systems that can learn to solve any complex problem without needing to be taught how", and it works "from the premise that AI needs to be general".[92] DeepMind publishes papers on AGI, e.g. "PathNet: Evolution Channels Gradient Descent in Super Neural Networks" (Fernando et al. 2017).

Lead institution: Google

Partner institutions: none

Type of institution: public corporation

Open-source: yes[93]

Lead country: UK

Military connection: unspecified
- Google has extensive defense contracts in the US,[94] but these appear to be unrelated to DeepMind

Partner countries: USA, Canada
- DeepMind is based in London and also has a team at Google headquarters in Mountain View, California (Shead 2017). It recently opened an office in Edmonton, Alberta.[95]

Stated goals: humanitarianism
- Their website presents a slogan "Solve intelligence. Use it to make the world a better place."; it describes "AI as a multiplier for human ingenuity" to solve problems like climate change and healthcare; and states "We believe that AI should ultimately belong to the world, in order to benefit the many and not the few".[96] Similarly, it writes AI will be "helping humanity tackle some of its greatest challenges, from climate change to delivering advanced healthcare".[97]

Engagement on safety: active
- DeepMind insisted on an AI ethics board at Google during its acquisition (Gibbs 2014). It collaborates with OpenAI on long-term AI safety projects.[98] It also participates independently from Google in the Partnership on AI to Benefit People & Society.

Size: large

[92] https://deepmind.com/blog/open-sourcing-deepmind-lab
[93] https://github.com/deepmind
[94] https://www.fpds.gov/ezsearch/fpdsportal?q=google+DEPARTMENT_FULL_NAME%3A"DEPT+OF+DEFENSE"
[95] https://deepmind.com/blog/deepmind-office-canada-edmonton
[96] https://deepmind.com/about
[97] https://deepmind.com/blog/learning-through-human-feedback
[98] https://blog.openai.com/deep-reinforcement-learning-from-human-preferences

DeSTIN (Deep SpatioTemporal Inference Network)

Main website: http://web.eecs.utk.edu/~itamar/Papers/BICA2009.pdf and http://wiki.opencog.org/w/DeSTIN

DeSTIN was initially developed by Itamar Arel and colleagues at the University of Tennessee. It is also being developed by the OpenCog open-source AI project. DeSTIN uses deep learning for pattern recognition. The OpenCog website states that OpenCog "has adopted this academic project to prepare it for open-source release".[99] Goertzel (2014, p.17) notes that DeSTIN "has been integrated into the CogPrime architecture… but is primarily being developed to serve as the center of its own AGI design".

Lead institution: University of Tennessee

Partner institution: OpenCog Foundation

Type of institution: academic, nonprofit

Open-source: yes[100]

Military connection: unspecified

Lead country: USA

Partner countries: none

Stated goals: unspecified

Engagement on safety: unspecified

Size: small

[99] http://wiki.opencog.org/w/DeSTIN
[100] https://github.com/opencog/destin

DSO-CA

Main website: none found

DSO-CA is a project of Gee Wah Ng and colleagues at DSO National Laboratories, which is Singapore's primary national defense research agency. It is "a top-level cognitive architecture that models the information processing in the human brain", with similarities to LIDA, CogPrime, and other AGI cognitive architectures.[101]

Lead institution: DSO National Laboratories

Partner institution: none

Type of institution: government
- DSO was "corporatized" in 1997[102] but is listed as a government agency on its LinkedIn page.[103]

Open-source: no

Military connection: yes

Lead country: Singapore

Partner countries: none

Stated goals: unspecified

Engagement on safety: unspecified

Size: small

[101] Ng et al. (2017)
[102] https://www.dso.org.sg/about/history
[103] https://www.linkedin.com/company-beta/15618

FLOWERS (FLOWing Epigenetic Robots and Systems)

Main website: https://flowers.inria.fr

FLOWERS is led by Pierre-Yves Oudeyer of Inria (Institut National de Recherche en Informatique et en Automatique, or French National Institute for Research in Computer Science and Automation) and David Filliat of ENSTA ParisTech. The project "studies mechanisms that can allow robots and humans to acquire autonomously and cumulatively repertoires of novel skills over extended periods of time".[104]

Lead institutions: Inria and ENSTA ParisTech

Partner institutions: none

Type of institution: academic, government
- Inria is a government research institute; ENSTA ParisTech is a public college

Open-source: yes[105]

Military connection: unspecified[106]

Lead country: France

Partner countries: none

Stated goals: intellectualist
- The project website focuses exclusively on intellectual aspects of its AI research and also cognitive science alongside AI as one of its two research strands.[107]

Engagement on safety: active
- FLOWERS has explored safety in the context of human-robot interactions.[108]

Size: medium-small

[104] https://flowers.inria.fr
[105] https://flowers.inria.fr/software
[106] Funding reported in recent publications comes mainly from government science foundations
[107] https://flowers.inria.fr
[108] Oudeyer et al. (2011)

GoodAI

Main website: https://www.goodai.com

GoodAI is a privately held corporation led by computer game entrepreneur Marek Rosa. It is based in Prague. It is funded by Rosa, who has invested at least $10 million in it. Rosa states that "GoodAI is building towards my lifelong dream to create general artificial intelligence. I've been focused on this goal since I was 15 years old".[109]

Lead institution: GoodAI

Partner institutions: none

- GoodAI lists several partner organizations, including one, SlovakStartup, based in Slovakia.[110] However, none of the partners are counted in this study because they do not contribute AGI R&D.

Type of institution: private corporation

Open-source: yes[111]

Military connection: unspecified

Lead country: Czech Republic

Partner countries: none

Stated goals: humanitarianism, intellectualist

- The GoodAI website states that its "mission is to develop general artificial intelligence - as fast as possible - to help humanity and understand the universe" and that it aims "to build general artificial intelligence that can find cures for diseases, invent things for people that would take much longer to invent without the cooperation of AI, and teach us much more than we currently know about the universe."[112] It emphasizes that building AGI "is not a race. It's not about competition, and not about making money.[113]

Engagement on safety: active

- GoodAI reports having a dedicated AI safety team[114] and cites Nick Bostrom's *Superintelligence* as a research inspiration on AI safety.[115]

Size: medium

[109] https://www.goodai.com/about
[110] https://www.goodai.com/partners
[111] https://github.com/GoodAI
[112] https://www.goodai.com/about
[113] https://www.goodai.com/about
[114] https://www.goodai.com/about
[115] https://www.goodai.com/research-inspirations

HTM (Hierarchical Temporal Memory)

Main website: https://numenta.com

HTM is developed by the Numenta corporation of Redwood City, California and an open-source community that the corporation hosts. HTM is led by Jeffrey Hawkins, who previously founded Palm Computing. HTM is based on a model of the human neocortex. Their website states, "We believe that understanding how the neocortex works is the fastest path to machine intelligence… Numenta is far ahead of any other team in this effort to create true machine intelligence".[116]

Lead institution: Numenta

Partner institutions: none

Type of institution: private corporation

Open-source: yes[117]
- Numenta offers a fee-based commercial license and an open-source license.[118]

Military connection: unspecified
- HTM was used in a 2008 Air Force Institute of Technology student thesis (Bonhoff 2008)

Lead country: USA

Partner countries: none

Stated goals: humanitarianism, intellectualist
- The Numenta corporate website lists two agendas: (1) neuroscience research, with an intellectualist theme, e.g. "Numenta is tackling one of the most important scientific challenges of all time: reverse engineering the neocortex", and (2) machine intelligence technology, with a humanitarian theme, e.g. stating it is "important for the continued success of humankind".[119]
- Hawkins writes that "The future success and even survival of humanity may depend on" humanity "building truly intelligent machines", citing applications in energy, medicine, and space travel.[120]
- The Numenta Twitter states that "Only if we make AI a public good, rather than the property of a privileged few, we can truly change the world."[121]

Engagement on safety: dismissive
- Hawkins has dismissed concerns about AGI as a catastrophic risk, stating "I don't see machine intelligence posing any threat to humanity" (Hawkins 2015).

Size: medium

[116] https://numenta.com
[117] https://numenta.org
[118] https://numenta.com/assets/pdf/apps/licensing-guide.pdf
[119] https://numenta.com
[120] Hawkins (2017) and the Numenta blog: https://numenta.com/blog/2017/06/IEEE-special-edition-article-by-Jeff
[121] https://twitter.com/Numenta/status/892445032736849920

Human Brain Project (HBP)

Main website: http://www.humanbrainproject.eu

HBP is a project for neuroscience research and brain simulation. It is sponsored by the European Commission, with a total of $1 billion committed over ten years beginning 2013.[122] Initially led by Henry Makram, it was reorganized following extended criticism (Theil 2015). It is based at École Polytechnique Fédérale de Lausanne, with collaborating institutions from around Europe and Israel.[123] It hosts platforms for brain simulation, neuromorphic computing, and neurorobotics.[124] Markram also founded Blue Brain, which shares research strategy with HBP.[125]

Reason for consideration: A large-scale brain research project, similar to Blue Brain

Lead institutions: École Polytechnique Fédérale de Lausanne

Partner institutions: 116 listed on the HBP website[126]

Type of institution: academic

Open-source: restricted
- Obtaining an account requires a request and sharing a copy of one's passport.[127]

Military connection: no
- HBP policy forbids military applications[128]

Lead country: Switzerland

Partner countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Hungary, Israel, Italy, Netherlands, Norway, Portugal, Slovenia, Spain, Sweden, Turkey, and United Kingdom

Stated goals: animal welfare, humanitarian, intellectualist
- HBP pursues brain simulation to "reduce the need for animal experiments" and "study diseases".[129] It also lists "understanding cognition" as a core theme.[130]

Engagement on safety: unspecified
- HBP has an ethics program focused on research procedure, not societal impacts.[131]

Size: large

[122] http://www.humanbrainproject.eu/en/science/overview/
[123] http://www.humanbrainproject.eu/en/open-ethical-engaged/contributors/partners
[124] http://www.humanbrainproject.eu/en/brain-simulation/brain-simulation-platform; http://www.humanbrainproject.eu/en/silicon-brains; http://www.humanbrainproject.eu/en/robots
[125] http://bluebrain.epfl.ch/page-52741-en.html
[126] http://www.humanbrainproject.eu/en/open-ethical-engaged/contributors/partners
[127] http://www.humanbrainproject.eu/en/silicon-brains/neuromorphic-computing-platform
[128] https://nip.humanbrainproject.eu/documentation, https://nip.humanbrainproject.eu/documentation
[129] http://www.humanbrainproject.eu/en/brain-simulation
[130] http://www.humanbrainproject.eu/en/understanding-cognition
[131] https://www.humanbrainproject.eu/en/open-ethical-engaged/ethics/ethics-management

Icarus

Main website: http://cll.stanford.edu/research/ongoing/icarus

Icarus is led by Pat Langley of Stanford University. Another active contributor is Dongkyu Choi of University of Kansas, a former Langley Ph.D. student. Icarus is a cognitive architecture project similar to ACT-R and Soar, emphasizing perception and action in physical environments.[132] Its website is out of date but several recent papers have been published.[133]

Lead institution: Stanford University

Partner institutions: University of Kansas

Type of institution: academic

Open-source: no

Military connection: yes[134]

Lead country: USA

Partner countries: none

Stated goals: intellectualist

- Choi and Langley (2017) write that "our main goal" is "achieving broad coverage of cognition functions" in "the construction of intelligent agents".

Engagement on safety: unspecified

Size: small

[132] Goertzel (2014); Choi and Langley (2017)
[133] See e.g. Menager and Choi (2016); Choi and Langley (2017); Langley (2017)
[134] The Icarus group reports funding from the US Office of Naval Research, Navy Research Lab, and DARPA in Choi and Langley (2017).

Leabra

Main website: https://grey.colorado.edu/emergent/index.php/Leabra

Leabra is led by Randall O'Reilly of University of Colorado. It is a cognitive architecture project emphasizing modeling neural activity. A recent paper states that Leabra is "a long-term effort to produce an internally consistent theory of the neural basis of human cognition", and that "More than perhaps any other proposed cognitive architecture, Leabra is based directly on the underlying biology of the brain, with a set of biologically realistic neural processing mechanisms at its core".[135] It was briefly connected to ACT-R via the SAL project.

Lead institution: University of Colorado

Partner institutions: none

Type of institution: academic

Open-source: yes[136]

Military connection: yes[137]

Lead country: USA

Partner countries: none

Stated goals: intellectualist

Engagement on safety: unspecified

Size: medium-small

[135] O'Reilly et al. (2016)
[136] https://grey.colorado.edu/emergent/index.php/Main_Page
[137] The Leabra group reports funding from the US Office of Naval Research and Army Research Lab at https://grey.colorado.edu/CompCogNeuro/index.php/CCNLab/funding

LIDA (Learning Intelligent Distribution Agent)

Main website: http://ccrg.cs.memphis.edu

LIDA is led by Stan Franklin of University of Memphis. It is based on Bernard Baars's Global Workspace Theory, "integrating various forms of memory and intelligent processing in a single processing loop" (Goertzel 2014, p.24). Goertzel (2014, p.24) states that LIDA has good grounding in neuroscience but is only capable at "lower level" intelligence, not more advanced thought. LIDA is supported by the US Office of Naval Research; a simpler version called IDA is being used to automate "the decision-making process for assigning sailors to their new posts".[138]

Lead institutions: University of Memphis

Partner institutions: none

Type of institution: academic

Open-source: restricted[139]
- Registration is required to download code; commercial use requires a commercial license

Military connection: yes

Lead country: USA

Partner countries: none

Stated goals: humanitarian, intellectualist
- The LIDA website says that it seeks "a full cognitive model of how minds work" and focuses predominantly on intellectual research aims.[140]
- A paper by LIDA researchers Tamas Madl and Stan Franklin states that robots need ethics "to constrain them to actions beneficial to humans".[141]
- Franklin hints at support for moral standing for AGI, noting the study of "synthetic emotions" in AI, which could suggest that an AGI "should be granted moral status".[142]

Engagement on safety: active
- The Madl and Franklin paper addresses AGI safety challenges like the subtlety of defining human ethics with the precision needed for programming.[143]
- Franklin has also collaborated with AI ethicists Colin Allen and Wendell Wallach on the challenge of getting AGIs to make correct moral decisions.[144]

Size: medium

[138] http://ccrg.cs.memphis.edu
[139] http://ccrg.cs.memphis.edu/framework.html
[140] http://ccrg.cs.memphis.edu
[141] Madl and Franklin (2015)
[142] Wallach et al. (2011, p.181)
[143] Madl and Franklin (2015)
[144] Wallach et al. (2010)

Maluuba

Main website: http://www.maluuba.com

Maluuba is an AI company based in Montreal, recently acquired by Microsoft. Maluuba writes: "our vision has been to solve artificial general intelligence by creating literate machines that could think, reason and communicate like humans".[145]

Lead institution: Microsoft

Partner institutions: none

Type of institution: public corporation

Open-source: yes[146]

Military connection: unspecified

Lead country: Canada

Partner countries: USA

Stated goals: intellectualist, profit

- Maluuba writes that "understanding human language is extremely complex and is ultimately the holy grail in the field of Artificial Intelligence", and that they aim "to solve fundamental problems in language understanding, with the vision of creating a truly literate machine".[147]
- Maluuba VP of Product Mo Musbah says that Maluuba wants general AI so that "it can scale in terms of how it applies in an AI fashion across different industries".[148]
- Microsoft is a founding partner of the Partnership on AI to Benefit People & Society, which has humanitarian goals,[149] but this does not appear to have transferred to Maluuba's goals.

Engagement on safety: moderate

- Maluuba describes safety as an important research challenge.[150] Maluuba researcher Harm van Seijen writes that "I think such discussions [about AI safety] are good, although we should be cautious of fear mongering."[151] Microsoft is also a founding partner of the Partnership on AI to Benefit People & Society, which expresses concern about AI safety.[152] No direct safety activity by Maluuba was identified.

Size: medium

[145] http://www.maluuba.com/blog/2017/1/13/maluuba-microsoft
[146] https://github.com/Maluuba
[147] http://www.maluuba.com/blog/2017/1/13/maluuba-microsoft
[148] Townsend (2016)
[149] https://www.partnershiponai.org/tenets
[150] http://www.maluuba.com/blog/2017/3/14/the-next-challenges-for-reinforcement-learning
[151] Townsend (2016)
[152] https://www.partnershiponai.org/tenets

MicroPsi

Main website: http://cognitive-ai.com

MicroPsi is led by Joscha Bach of the Harvard Program for Evolutionary Dynamics. Bach's mission is reportedly "to build a model of the mind is the bedrock research in the creation of Strong AI, i.e. cognition on par with that of a human being".[153] Goertzel (2014, p.24) states that MicroPsi has good grounding in neuroscience and is only capable at lower level intelligence. A recent paper on MircoPsi says that it is "an architecture for Artificial General Intelligence, based on a framework for creating and simulating cognitive agents", which began in 2003.[154]

Lead institutions: Harvard University

Partner institutions: none

Type of institution: academic

Open-source: yes[155]

Military connection: unspecified

Lead country: USA

Partner countries: none

Stated goals: intellectualist

- The MicroPsi website says that "MicroPsi is a small step towards understanding how the mind works".[156]

Engagement on safety: unspecified

Size: small

[153] http://bigthink.com/experts/joscha-bach
[154] Bach (2015)
[155] https://github.com/joschabach/micropsi2
[156] http://cognitive-ai.com

Microsoft Research AI (MSR AI)

Main website: https://www.microsoft.com/en-us/research/lab/microsoft-research-ai

MSR AI is an AI "research and incubation hub" at Microsoft announced in July 2017.[157] The project seeks "to probe the foundational principles of intelligence, including efforts to unravel the mysteries of human intellect, and use this knowledge to develop a more general, flexible artificial intelligence".[158] The project pulls together more than 100 researchers from different branches of AI at Microsoft's Redmond headquarters.[159] By pulling together different branches, MSR AI hopes to achieve more sophisticated AI, such as "systems that understand language and take action based on that understanding".[160] However, it has also been criticized for a potentially unwieldy organizational structure.[161]

Lead institutions: Microsoft

Partner institutions: none

Type of institution: public corporation

Open-source: no

Military connection: unspecified
- Microsoft has a Military Affairs program,[162] but its link to MSR AI is unclear.

Lead country: USA

Partner countries: none

Stated goals: humanitarian, intellectualist
- Microsoft CEO Satya Nadella states broadly humanitarian goals, such as "AI must be designed to assist humanity".[163]
- The MSR AI website states that it aims "to solve some of the toughest challenges in AI" and "probe the foundational principles of intelligence".[164]

Engagement on safety: unspecified
- The MSR AI group on Aerial Informatics and Robotics has extensive attention to safety, but this is for the narrow context of aircraft, not for AGI.[165]

Size: medium-large

[157] https://blogs.microsoft.com/blog/2017/07/12/microsofts-role-intersection-ai-people-society
[158] https://www.microsoft.com/en-us/research/lab/microsoft-research-ai
[159] Etherington (2017)
[160] https://blogs.microsoft.com/blog/2017/07/12/microsofts-role-intersection-ai-people-society
[161] Architecht (no date)
[162] https://military.microsoft.com/about
[163] Nadella (2016)
[164] https://www.microsoft.com/en-us/research/lab/microsoft-research-ai
[165] https://www.microsoft.com/en-us/research/group/air

MLECOG

Main website: none

MLECOG is a cognitive architecture project led by Janusz Starzyk of Ohio University. A paper on MLECOG describes it as similar to NARS and Soar.[166] MLECOG is an acronym for Motivated Learning Embodied Cognitive Architecture.

Lead institution: Ohio University

Partner institutions: none

Type of institution: academic

Open-source: no

Military connection: unspecified

Lead country: USA

Partner countries: none

Stated goals: unspecified

Engagement on safety: unspecified

Size: small

[166] Starzyk and Graham (2015)

NARS

Main website: https://sites.google.com/site/narswang

NARS is an AGI research project led by Pei Wang of Temple University. NARS is an acronym for Non-Axiomatic Reasoning System, in reference to the AI being based on tentative experience and not axiomatic logic, consistent with its "assumption of insufficient knowledge and resources".[167] In a 2011 interview, Wang suggests that NARS may achieve human-level AI by 2021.[168]

Lead institutions: Temple University

Partner institutions: none

Type of institution: academic

Open-source: yes[169]

Military connection: unspecified

Lead country: USA

Partner countries: none

Stated goals: humanitarian, intellectualist

- The NARS website explains that NARS is "morally neutral" in the sense that it can be programmed with any moral system.[170]
- The NARS website emphasizes that NARS should aim for a positive impact "on the human society" and be "human-friendly".[171]
- The NARS website states that "the ultimate goal of this research is to fully understand the mind, as well as to build thinking machines".[172]

Engagement on safety: active

- Wang has written on safety issues in NARS, such as "motivation management", a factor in the ability of NARS to reliably pursue its goals and not be out of control.[173]

Size: medium

[167] https://sites.google.com/site/narswang/home/nars-introduction
[168] Goertzel (2011)
[169] https://github.com/opennars
[170] https://sites.google.com/site/narswang/EBook/Chapter5/section-5-5-education
[171] https://sites.google.com/site/narswang/EBook/Chapter5/section-5-5-education
[172] https://sites.google.com/site/narswang
[173] Wang (2012)

Nigel

Main website: http://kimera.ai

Nigel is the AGI project of Kimera, an AI corporation based in Portland, Oregon. Kimera was founded in 2005 by Mounir Shita and Nicholas Gilman. It styles itself as "The AGI Company".[174] In 2016, Kimera unveiled Nigel, which it claims is "the first commercially deployable artificial general intelligence technology".[175] However, as of 2016, little about Nigel is publicly available and critics are skeptical about the AGI claim.[176] Nigel has been described as a personal assistant bot similar to Apple's SIRI and Amazon's Alexa.[177] Kimera also envisions Nigel being used for a variety of online activities, "bringing about a greater transformation in global business than even the internet", and transforming "the internet from a passive system of interconnections into a proactive, intelligent global network".[178]

Lead institution: Kimera

Partner institutions: none

Type of institution: private corporation

Open-source: no

Military connection: unspecified

Lead country: USA

Partner countries: none

Stated goals: humanitarianism

- Kimera presents a humanitarian vision for AGI, writing that "Artificial General Intelligence has the power to solve some - or all - of humanity's biggest problems, such as curing cancer or eliminating global poverty."[179]

Engagement on safety: unspecified

Size: medium-small

[174] http://kimera.ai
[175] http://kimera.ai
[176] Boyle (2016); Jee (2016)
[177] Boyle (2016)
[178] http://kimera.ai/nigel
[179] http://kimera.ai/company

NNAISENSE

Main website: https://nnaisense.com

NNAISENSE is a private company based in Lugano, Switzerland. Several of its team members have ties to the Dalle Molle Institute for Artificial Intelligence (IDSIA, a Swiss nonprofit research institute), including co-founder and Chief Scientist Jürgen Schmidhuber. Its website states that it seeks "to build large-scale neural network solutions for superhuman perception and intelligent automation, with the ultimate goal of marketing general-purpose Artificial Intelligences."[180]

Lead institutions: NNAISENSE

Partner institutions: none

Type of institution: private company

Open-source: no

Military connection: unspecified

Lead country: Switzerland

Partner countries: none

Stated goals: intellectualism, profit

- Schmidhuber is described as a "consummate academic" who founded the company to prevent other companies from poaching his research talent; NNAISENSE reportedly "chooses projects based on whether they'll benefit the machine's knowledge, not which will bring in the highest fees".[181]
- The NNAISENSE website states "the ultimate goal of *marketing*" AGI (emphasis added).[182]

Engagement on safety: unspecified

Size: medium-small

[180] https://nnaisense.com
[181] Webb (2017)
[182] https://nnaisense.com

OpenAI

Main website: https://openai.com

OpenAI is a nonprofit AI research organization founded by several prominent technology investors. It is based in San Francisco. Its funders have pledged $1 billion to the project. Its website states: "Artificial general intelligence (AGI) will be the most significant technology ever created by humans. OpenAI's mission is to build safe AGI, and ensure AGI's benefits are as widely and evenly distributed as possible."[183] It is part of the Partnership on AI to Benefit People & Society.[184]

Lead institution: OpenAI

Partner institutions: none

Type of institution: nonprofit

Open-source: yes[185]

Military connection: unspecified

Lead country: USA

Partner countries: none

Stated goals: humanitarianism
- OpenAI seeks that AGI "leads to a good outcome for humans,"[186] and that "AGI's benefits are as widely and evenly distributed as possible."[187]

Engagement on safety: active
- Safe AGI is part of OpenAI's mission. While it releases much of its work openly, its website states that "in the long term, we expect to create formal processes for keeping technologies private when there are safety concerns".[188] It also collaborates with DeepMind on long-term AI safety.[189]

Size: large

[183] https://openai.com/about
[184] https://www.partnershiponai.org/partners
[185] https://github.com/openai
[186] https://openai.com/jobs
[187] https://openai.com/about
[188] https://openai.com/about
[189] https://blog.openai.com/deep-reinforcement-learning-from-human-preferences

Real AI

Main website: http://realai.org

Real AI is a private company in Hong Kong led by Jonathan Yan. It is a single member company.[190] Its mission is "to ensure that humanity has a bright future with safe AGI".[191] It works on strategy for safe AGI and technical research in deep learning, the latter on the premise that deep learning can scale up to AGI.[192] Its website states that "We align ourselves with effective altruism and aim to benefit others as much as possible".[193]

Lead institution: Real AI

Partner institutions: none

Type of institution: private corporation

Open-source: no

Military connection: unspecified

Lead country: China

Partner countries: none

Stated goals: humanitarian

Engagement on safety: active

- Real AI has a dedicated page surveying ideas about AGI safety[194] and an extended discussion of its own thinking.[195]

Size: small

[190] http://realai.org/about/admin
[191] http://realai.org/about
[192] http://realai.org/prosaic
[193] http://realai.org/about
[194] http://realai.org/safety
[195] http://realai.org/blog/towards-safe-and-beneficial-intelligence

Research Center for Brain-Inspired Intelligence (RCBII)

Main website: http://bii.ia.ac.cn

RCBII is a "long term strategic scientific program proposed by Institute of Automation, Chinese Academy of Sciences".[196] The group is based in Beijing.[197] RCBII does research in fundamental neuroscience, brain simulation, and AI. It states that "Brain-inspired Intelligence is the grand challenge for achieving Human-level Artificial Intelligence".[198]

Lead institution: Chinese Academy of Sciences

Partner institutions: none

Type of institution: government

- The Chinese Academy of Sciences is a public institution under the Chinese government[199]

Open-source: no

Military connection: unspecified

Lead country: China

Partner countries: none

Stated goals: intellectualist

- The RCBII website only lists intellectual motivations, stating "The efforts on Brain-inspired Intelligence focus on understanding and simulating the cognitive brain at multiple scales as well as its applications to brain-inspired intelligent systems." No social or ethical aspects of these applications are discussed.

Engagement on safety: unspecified

Size: medium-small

[196] http://bii.ia.ac.cn/about.htm
[197] http://english.ia.cas.cn/au/fu/
[198] http://bii.ia.ac.cn/about.htm
[199] http://english.cas.cn/about_us/introduction/201501/t20150114_135284.shtml

Sigma

Main website: http://cogarch.ict.usc.edu

Sigma is led by Paul Rosenbloom of University of Southern California. It has a publication record dating to 2009[200] and won awards at the 2011 and 2012 AGI conferences.[201] Rosenbloom was previously a co-PI of Soar.[202]

Lead institution: University of Southern California

Partner institutions: none

Type of institution: academic

Open-source: yes[203]

Military connection: yes[204]

Lead country: USA

Partner countries: none

Stated goals: intellectualist

- The Sigma website says its goal is "to develop a sufficiently efficient, functionally elegant, generically cognitive, grand unified, cognitive architecture in support of virtual humans (and hopefully intelligent agents/robots – and even a new form of unified theory of human cognition – as well)."[205]
- Rosenbloom also hints at cosmist views in a 2013 interview, stating "I see no real long-term choice but to define, and take, the ethical high ground, even if it opens up the possibility that we are eventually superseded – or blended out of pure existence – in some essential manner."[206]

Engagement on safety: unspecified

- In a 2013, interview, Rosenbloom hints at being dismissive, questioning "whether superhuman general intelligence is even possible", but also explores some consequences if it is possible, all while noting his lack of "any particular expertise" on the matter.[207] No indications of safety work for Sigma were found.

Size: medium

[200] http://cogarch.ict.usc.edu/publications-new
[201] http://cs.usc.edu/~rosenblo
[202] http://cs.usc.edu/~rosenblo
[203] https://bitbucket.org/sigma-development/sigma-release/wiki/Home
[204] Funding from the US Army, Air Force Office of Scientific Research, and Office of Naval Research is reported in Rosenbloom (2013)
[205] http://cogarch.ict.usc.edu
[206] https://intelligence.org/2013/09/25/paul-rosenbloom-interview
[207] https://intelligence.org/2013/09/25/paul-rosenbloom-interview

SiMA

Main website: http://sima.ict.tuwien.ac.at/description

SiMA is a project led by Dietmar Dietrich of Vienna University of Technology. SiMA is an acronym for Simulation of the Mental Apparatus & Applications. The project aims "to develop a broad human-like intelligent system that is able to cope with complex and dynamic problems rather than with narrowly and well-defined domains".[208] It includes extensive attention to psychoanalysis, especially Freud and other German-language scholars. It was started by Dietrich in 1999.[209]

Lead institutions: Vienna University of Technology

Partner institutions: none

Type of institution: academic

Open-source: yes[210]
- The open-source portion of the project appears to be out of date

Military connection: unspecified

Lead country: Austria

Partner countries: none
- A project document states collaborators in Canada, Portugal, South Africa, and Spain, but details of these collaborations could not be identified.[211]

Stated goals: intellectualist
- A project document states that SiMA was founded "to understand how the brain as a whole works".[212]
- The document also discusses applications in automation and the prospect that "machines will have feelings",[213] but no specific goals could be identified from this.

Engagement on safety: unspecified

Size: medium

[208] http://sima.ict.tuwien.ac.at/description
[209] Brandstätter et al. (2015, p.V)
[210] http://sima.ict.tuwien.ac.at/wiki/index.php/Main_Page
[211] Brandstätter et al. (2015, p.IV)
[212] Brandstätter et al. (2015, p.V)
[213] Brandstätter et al. (2015, p.XI)

SingularityNET

Main website: https://singularitynet.io

SingularityNET is an AGI project led by Ben Goertzel. It was publicly launched in 2017.[214] It aims to be a platform in which anyone can post AI code or use AI code posted by other people. It plans to use cryptocurrency for payments for the use of AI on its site. This setup aims to makie AI more democratic than what could occur via governments or corporations (Goertzel 2017b). SingularityNET plans to have decision making done by voting within its user community (Goertzel et al. 2017).

Lead institutions: SingularityNET Foundation

Partner institutions: OpenCog Foundation
- Several other partners are listed on the SingularityNET website, but OpenCog Foundation is the only one that contributes AGI R&D.

Type of institution: nonprofit[215]

Open-source: yes[216]

Military connection: unspecified

Lead country: China[217]

Partner countries: Australia, Brazil, Canada, Germany, Portugal, Russia, USA[218]

Stated goals: animal welfare, ecocentric, humanitarian, transhumanist
- SingularityNET is described as "for the People (and the Robots!)" and "the happiness of sentient beings", with "benefits for all people, and for all life" (Goertzel 2017b; Goertzel et al. 2017).
- SingularityNET also states a goal of profit, but describes this as a means to other goals, for example stating "SingularityNET has the potential to profit tremendously… [and] to direct the profit thus generated to apply AI for global good" (Goertzel 2017b).

Engagement on safety: unspecified

Size: medium-small

[214] Blog posts at https://blog.singularitynet.io date to October 2017.
[215] SingularityNET Foundation is described as a nonprofit on p.8 of Goertzel et al. (2017).
[216] https://github.com/singnet/singnet
[217] CEO Goertzel and Chairman David Hanson are both based in Hong Kong (https://twitter.com/bengoertzel; https://twitter.com/hansonrobotics), which was also the site of a developer gathering: https://blog.singularitynet.io/singularitynet-tech-development-update-v1-0-4de5f87b4f42
[218] These are the countries that could be identified for SingularityNET personnel listed at https://singularitynet.io: Australia (Sergei Sergienko, https://www.linkedin.com/in/sergeisergienko); Brazil (Cassio Pennachin, http://www.pennachin.com); Canada (Tal Ball, https://www.linkedin.com/in/talgball); Germany (Trent McConaghy, https://www.linkedin.com/in/trentmc); Portugal (Mitchell Loureiro, https://www.linkedin.com/in/mitchelllloureiro); Russia (Alexey Potapov, Potapov et al. 2016; Anton Kolonin, Russia: https://www.linkedin.com/in/anton-kolonin-81a47836); and USA (Eddie Monroe, https://www.linkedin.com/in/eddiemonroe; Linas Vepstas, https://linas.org/resume.html; Jim Rutt, https://en.wikipedia.org/wiki/Jim_Rutt)

SNePS (Semantic Network Processing System)

Main website: http://www.cse.buffalo.edu/sneps

SNePS is led by Stuart Shapiro at State University of New York at Buffalo, with a publication record dating to 1969.[219] According to its website, its long-term goal is "to understand the nature of intelligent cognitive processes by developing and experimenting with computational cognitive agents that are able to use and understand natural language, reason, act, and solve problems in a wide variety of domains".[220]

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Lead institutions: State University of New York at Buffalo

Partner institutions: none

Type of institution: academic

Open-source: yes[221]

Military connection: yes[222]

Lead country: USA

Partner countries: none

Stated goals: intellectualist

Engagement on safety: unspecified

Size: medium

[219] http://www.cse.buffalo.edu/sneps/Bibliography
[220] http://www.cse.buffalo.edu/sneps
[221] https://github.com/SNePS/CSNePS, https://www.cse.buffalo.edu/sneps/Downloads
[222] US Army Research Office funding is reported in recent papers including Shapiro and Schlegel (2016).

Soar

Main website: http://soar.eecs.umich.edu and https://soartech.com

Soar is led by John Laird of University of Michigan and a spinoff corporation SoarTech, also based in Ann Arbor, Michigan. Laird and colleagues began Soar in 1981.[223] SOAR is an acronym for State, Operator Apply Result.

Lead institution: University of Michigan, SoarTech

Partner institutions: Pace University, Pennsylvania State University, University of Portland, and University of Southern California in the United States, University of Portsmouth in the United Kingdom, and Bar Ilan University and Cogniteam (a privately held company) in Israel.[224]

Type of institution: academic, private corporation

Open-source: yes[225]

Military connection: yes

- SoarTech lists customers including research laboratories of the US Air Force, Army, and Navy, and the US Department of Transportation.[226]

Lead country: USA

Partner countries: Israel, UK

Stated goals: intellectualist

- The Soar website describes it as an investigation into "an approximation of complete rationality" aimed at having "all of the primitive capabilities necessary to realize the complete suite of cognitive capabilities used by humans".[227]

Engagement on safety: unspecified

Size: medium-large

[223] http://ai.eecs.umich.edu/people/laird
[224] https://soar.eecs.umich.edu/groups
[225] https://github.com/SoarGroup, https://soar.eecs.umich.edu/Downloads
[226] http://soartech.com/about
[227] http://soar.eecs.umich.edu

Susaro

Main website: http://www.susaro.com

Susaro is an AI corporation based in the Cambridge, UK area. Its website states that it is in stealth mode, and that it is "designing the world's most advanced *artificial general intelligence* systems" (emphasis original) using an approach that "is a radical departure from conventional AI".[228] The Susaro website does not list personnel, but external websites indicate that it is led by AGI researcher Richard Loosemore.[229]

Lead institution: Susaro

Partner institutions: none

Type of institution: private corporation

Open-source: no
- Susaro has a GitHub page with no content[230]

Military connection: unspecified

Lead country: UK

Partner countries: none

Stated goals: ecocentric, humanitarian
- The Susaro website states that it aims to advance "human and planetary welfare… without making humans redundant".[231]

Engagement on safety: active
- The Susaro website states that "the systems we build will have an unprecedented degree of safety built into them… making it virtually impossible for them to become unfriendly".[232]

Size: unspecified

[228] All info and quotes are from http://www.susaro.com
[229] https://www.linkedin.com/in/richard-loosemore-47a2164, https://www.researchgate.net/profile/Richard_Loosemore, https://cofounderslab.com/profile/richard-loosemore
[230] https://github.com/susaroltd
[231] All info and quotes are from http://www.susaro.com
[232] All info and quotes are from http://www.susaro.com

Tencent AI Lab (TAIL)

Main website: http://ai.tencent.com/ailab

TAIL is the AI group of Tencent, the Shenzhen-based Chinese technology company. Its website lists several research areas, one of which is machine learning, which it says includes "general AI".[233] TAIL director Tong Zhang writes that TAIL "not only advances the state of art in artificial general intelligence, but also supports company products".[234]

Lead institution: Tencent

Partner institutions: none

Type of institution: public corporation
- Its website states that "Tencent will open-source its AI solutions in the areas of image, voice, security to its partners through Tencent Cloud", but it does not state that its AGI research is open-source.[235]

Open-source: no
- Tencent releases some work open-source,[236] but not its AGI

Military connection: unspecified

Lead country: China

Partner countries: USA
- TAIL recently opened an office in Seattle.[237]

Stated goals: unspecified

Engagement on safety: unspecified

Size: medium-small

[233] http://ai.tencent.com/ailab
[234] http://tongzhang-ml.org/research.html
[235] http://ai.tencent.com/ailab
[236] https://github.com/Tencent
[237] Mannes (2017)

Uber AI Labs (UAIL)

Main website: https://www.uber.com/info/ailabs

UAIL is the AI research division of Uber. UAIL began in 2016 with the acquisition of Geometric Intelligence (Temperton 2016), a private company founded in 2014 by Gary Marcus, Kenneth Stanley, and Zoubin Ghahramani in 2014 with incubation support from NYU.[238] Geometric Intelligence was based on Marcus's ideas for AGI, especially how to "learn with less training data" than is needed for deep learning (Chen 2017). Marcus has since left UAIL (Chen 2017). Upon acquisition by Uber, Geometric Ingelligence's personnel relocated to San Francisco, except for Ghahramani, who remained in Cambridge, UK (Metz 2016). UAIL is reportedly part of Uber's attempt to expand substantially beyond the private taxi market, similar to how Amazon expanded beyond books (Metz 2016). According to Ghahramani, the AI combines "some of the ideas in ruled-based learning with ideas in statistical learning and deep learning" (Metz 2016).

Lead institutions: Uber

Partner institutions: none

Type of institution: private corporation

Open-source: no
- Uber does have some open-source AI,[239] but this does not appear to include its AGI

Military connection: unspecified
- Uber does not appear to have any defense contracts.[240]

Lead country: USA

Partner countries: UK

Stated goals: humanitarian
- The UAIL website states that it seeks "to improve the lives of millions of people worldwide".[241]

Engagement on safety: unspecified
- No discussion of UAIL safety was found, except about the safety of Uber vehicles.[242]
- Marcus has indicated support for AI ethics research as long as it is understood that advanced AGI is not imminent.[243]

Size: medium

[238] https://www.nyu.edu/about/news-publications/news/2016/december/nyu-incubated-start-up-geometric-intelligence-acquired-by-uber.html
[239] https://github.com/uber/tensorflow
[240] https://www.fpds.gov/ezsearch/fpdsportal?q=uber+DEPARTMENT_FULL_NAME%3A"DEPT+OF+DEFENSE"
[241] https://www.uber.com/info/ailabs
[242] Chamberlain (2016)
[243] https://techcrunch.com/2017/04/01/discussing-the-limits-of-artificial-intelligence

Vicarious

Main website: https://www.vicarious.com

Vicarious is a privately held AI corporation founded in 2010 by Scott Phoenix and Dileep George and based in San Francisco. It has raised tens of millions of dollars in investments from several prominent investors.[244] Its states that it is "building systems to bring human-like intelligence to the world of robots".[245] In an interview, Phoenix says that Vicarious is working towards AGI, with "plenty of value created in the interim",[246] and that AGI would be "virtually the last invention humankind will ever make".[247]

Lead institution: Vicarious

Partner institutions: none

Type of institution: private corporation

Open-source: yes[248]

Military connection: unspecified

Lead country: USA

Partner countries: none

Stated goals: humanitarianism

- Vicarious is a Flexible Purpose Corporation, reportedly so that it can "pursue the maximization of social benefit as opposed to profit".[249] Scott Phoenix says that Vicarious aims to build AI "to help humanity thrive".[250]

Engagement on safety: moderate

- Phoenix says that AI safety may be needed "At some time in the future… but the research is at a really early stage now,"[251] and that it will not be difficult because AI will probably be "smart enough to figure out what it was that we wanted it to do."[252]

Size: medium-large

[244] Cutler (2014); High (2016)
[245] https://www.vicarious.com/research.html
[246] High (2016)
[247] TWiStartups (2016)
[248] https://github.com/vicariousinc
[249] High (2016)
[250] High (2016)
[251] Best (2016)
[252] TWiStartups (2016)

Victor

Main website: http://2ai.org/victor

Victor is the main project of 2AI, which is a subsidiary of the private company Cifer Inc.[253] 2AI is led by Timothy Barber and Mark Changizi and lists addresses in Boise, Idaho and the US Virgin Islands.[254] 2AI describes Victor as an AGI project.[255] Its website states that "we believe the future of AI will ultimately hinge upon its capacity for competitive interaction".[256]

Lead institutions: Cifer

Partner institutions: none

Type of institution: private corporation

Open-source: no

Military connection: unspecified

Lead country: USA

Partner countries: none

Stated goals: ecocentric

- Its website states that "2AI is a strong advocate of solutions for protecting and preserving aquatic ecosystems, particularly reefs and seamounts, which are crucial nexus points of Earth's biomass and biodiversity".[257]

Engagement on safety: dismissive

- The project website states that AI catastrophe scenarios are "crazy talk" because AI will need humans to maintain the physical devices it exists on and thus will act to ensure humans maintain and expand these devices.[258]

Size: small

[253] http://2ai.org/legal
[254] http://2ai.org/legal
[255] http://2ai.org/landscape
[256] http://2ai.org/victor
[257] http://2ai.org/legal
[258] http://www.2ai.org/killerai

Whole Brain Architecture Initiative (WBAI)

Main website: https://wba-initiative.org

WBAI is a nonprofit organization in Tokyo led by Hiroshi Yamakawa. Yamakawa is Director of AI at Dwango and also affiliated with Tamagawa University and the Japanese Society for Artificial Intelligence. WBAI's mission is "to create (engineer) a human-like artificial general intelligence (AGI) by learning from the architecture of the entire brain"[259]. They "aim for the construction of artificial general intelligence (AGI) to surpass the human brain capability around the year 2030".[260] WBAI receives support from, among others, Panasonic, Toshiba, and Toyota.[261]

Lead institution: Whole Brain Architecture Initiative

Partner institutions: none

Type of institution: nonprofit

Open-source: yes[262]

Military connection: unspecified

Lead country: Japan

Partner countries: none

Stated goals: humanitarianism

- WBAI promotes AI development that is "best for the human society".[263] Additionally, in a slideshow about WBAI, they quote Yamakawa as stating that "the grace and wealth that EcSIA [ecosystem of shared intelligent agents] affords needs to be properly distributed to everyone".[264]

Engagement on safety: active

- Safety is a significant theme for WBAI. For example, their website states, "In quantitative aspects of intelligence, AGI will overwhelm human beings. If AGI is heterogeneous to us, it may be difficult for us to understand, to embed ethics as a member of society, and to maintain control. Thus, we could say it is a relatively safe choice to build the first AGI in a form similar to us."[265] The implication here is that WBAI seeks to build brain-like AGI in part because that would be safer.

Size: medium-small

[259] https://wba-initiative.org/en/about/greetings
[260] https://wba-initiative.org/en/wba
[261] https://wba-initiative.org/en/supporting-members
[262] https://github.com/wbap
[263] https://wba-initiative.org/en/about/vision
[264] https://www.slideshare.net/HiroshiYamakawa/2017-0512gatsby-wsv10b-75941913
[265] https://wba-initiative.org/en/2071

Appendix 2. Other Notable Projects

4CAPS

Main website: http://www.ccbi.cmu.edu/projects_4caps.html

4CAPS was led by psychologist Marcel Just of Carnegie Mellon University. It is "a hybrid of a computational neuroscience model and a symbolic AI system" (Goertzel 2014, p.19). It "can account for both traditional behavioral data and, more interestingly, the results of neuroimaging studies".[266] The project reports funding by the Office of Naval Research and the Multidisciplinary Research Program of the University Research Initiative.[267]

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Apparently inactive

4D/RCS (Real-time Control Systems Architecture)

Main website: https://www.nist.gov/intelligent-systems-division/rcs-real-time-control-systems-architecture

4D/RCS was led by James Albus at the US National Institute of Standards and Technology. It consists of "hard-wired architecture and algorithms… augmented by learning" (Goertzel 2014, p.24).

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Apparently inactive

Achler

Main website: None

This unnamed project by Tsvi Achler of Los Alamos National Labs used neural networks in "a novel approach to bridging the symbolic-subsymbolic gap" Goertzel (2014, p.17).

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Apparently inactive

AGINAO

Main website: http://aginao.com

[266] http://www.ccbi.cmu.edu/projects_4caps.html
[267] http://www.ccbi.cmu.edu/projects_4caps.html

AGINAO was a project of Wojciech Skaba of Gdansk, Poland. It was active during 2011-2013,[268] and shows no activity more recently.

Reason for consideration: An AGI R&D project

Reason for exclusion: Apparently inactive

Alibaba

Main website: None

Alibaba is active in AI, but no indications of AGI were found.

Reason for consideration: Alibaba is a major computing technology company

Reason for exclusion: No indications of AGI projects were found

Amazon

Main website: https://aws.amazon.com/amazon-ai

Amazon has an AI group within its Amazon Web Services (AWS) division, but it does not appear to work on AGI. Amazon has donated AWS resources to OpenAI.[269]

Reason for consideration: Amazon is a major computing technology company

Reason for exclusion: No indications of AGI projects were found

Apple

Main website: https://machinelearning.apple.com

Apple has an AI group that does not appear to work on AGI. However, Apple has a reputation for secrecy and has only a minimal website. Apple is said to have less capable AI than companies like Google and Microsoft because Apple has stricter privacy rules, denying itself the data used to train AI.[270] Likewise, at least some of its AI research may be oriented towards learning from limited data or synthetic data.[271] Its recent AI company acquisitions are for narrow AI.[272] While it may be possible that Apple is working on AGI, no indications of this were found in online searches.

Reason for consideration: Apple is a major computing technology company

Reason for exclusion: No indications of AGI projects were found

[268] For example, Skaba (2012a; 2012b); http://aginao.com/page2.php
[269] https://blog.openai.com/infrastructure-for-deep-learning
[270] Vanian (2017a)
[271] Vanian (2017b)
[272] Tamturk (2017)

Artificial Brain Laboratory

Main website: None

The Artificial Brain Laboratory was led by Hugo de Garis at Xiamen University. The project appears to have ended upon de Garis's retirement around 2010. Xiamen University now has a Brain-like Intelligent Robotic Systems Group,[273] but this is not necessarily related to the ABL.

Reason for consideration: Author's prior knowledge

Reason for exclusion: Apparently inactive

Brain Imaging and Modeling Section (BIMS)

Main website: https://www.nidcd.nih.gov/research/labs/brain-imaging-and-modeling-section

BIMS is a research project led by Barry Horwitz of the US National Institutes of Health. The project combines brain imaging with computer modeling to advance basic neuroscience and treatment of brain disorders.

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Project focused on basic neuroscience, not the development of an AGI

BRAIN Initiative

Main website: https://www.braininitiative.nih.gov

BRAIN Initiative is a research project aimed at understanding the human brain. BRAIN is an acronym for Brain Research through Advancing Innovative Neurotechnologies. The project is based at the US National Institutes of Health and partners with several other US government agencies and private organizations.[274] Its website states that "By accelerating the development and application of innovative technologies, researchers will be able to produce a revolutionary new dynamic picture of the brain that, for the first time, shows how individual cells and complex neural circuits interact in both time and space."[275]

Reason for consideration: A large-scale brain research project, similar to Blue Brain

Reason for exclusion: Project focused on basic neuroscience, not the development of an AGI

Brain/MINDS

[273] http://information.xmu.edu.cn/en/?mod=departments&id=31
[274] https://www.braininitiative.nih.gov/about/index.htm
[275] https://www.braininitiative.nih.gov

Main website: http://brainminds.jp/en

Brain/MINDS is a neuroscience research project. Brain/MINDS is an acronym for Brain Mapping by Integrated Neurotechnologies for Disease Studies. The project is sponsored by Japan's Ministry of Education, Culture, Sports, Science, and Technology (MEXT). It focuses on the study of nonhuman primate brains, neural networks of brain disorders, and improving cooperation between basic and clinical neuroscience.[276]

Reason for consideration: A large-scale brain research project, similar to Blue Brain

Reason for exclusion: Project focused on basic neuroscience, not the development of an AGI

Carboncopies

Main website: https://www.carboncopies.org

Carboncopies is a nonprofit based in San Francisco "provides support to scientists in fields related to Whole brain emulation".[277]

Reason for consideration: A research project focused on technical details of AGI

Reason for exclusion: Project focused supporting AGI R&D, not on conducting R&D

CERA-CRANIUM

Main website: https://www.carboncopies.org

CERA-CRANIUM was a cognitive architecture project sometimes discussed in the context of AGI.[278] It was led by Raúl Arrabales of University of Madrid. It was used for computer games, winning a competition in 2010.[279] The most recent work on CERA-CRANIUM identified is from 2013.[280]

Reason for consideration: An AGI R&D project

Reason for exclusion: Apparently inactive

CHAI (Center for Human-Compatible AI)

Main website: http://humancompatible.ai

CHAI is a research group based at University of California, Berkeley. Its website states that its goal is "to develop the conceptual and technical wherewithal to reorient the general thrust of AI research towards provably beneficial systems", especially in the context of "machines that are more capable

[276] http://brainminds.jp/en/overview/greeting
[277] https://www.carboncopies.org/mission
[278] For example, Ng et al. (2017)
[279] Arrabales and Muñoz (2010)
[280] Arrabales et al. (2013)

than humans across a wide range of objectives and environments", which it sees as likely to eventually exist.[281]

Reason for consideration: A research project focused on technical details of AGI

Reason for exclusion: Project focused on safety aspects of AGI, not on the development of an AGI

CHREST

Main website: http://chrest.info

CHREST is led by Fernand Gobet of University of Liverpool. It was begun by Gobet in 1992 and traces to the 1959 EPAM system.[282] CHREST is an acronym for Chunk Hierarchy and REtrieval STructures. It is "a cognitive architecture that models human perception, learning, memory, and problem solving".[283] A paper on CHREST describes its strengths in categorization and understanding as being complementary to the strengths of other projects (e.g., ACT-R, Soar) in problem solving.[284] Its website lists no publications since 2010[285] and no updates since 2012.[286]

Reason for consideration: An AGI R&D project

Reason for exclusion: Apparently inactive

Cognitive Computing Project

Main website: http://research.ibm.com/cognitive-computing, especially https://www.research.ibm.com/cognitive-computing/neurosynaptic-chips.shtml

CCP is part of a suite of IBM AI projects, which also includes the famed Watson system. Goertzel (2014) discusses a project led by Dharmendra Modha to build computer hardware and software systems modeled after the human brain. The project has produced a new programming language and a new computer chip called TrueNorth, which Modha postulates as a "turning point in the history of computing".[287] The chip was introduced in a 2014 article in *Science*.[288] The chip development was supported by the DARPA SyNAPSE program aimed at making "low-power electronic neuromorphic computers that scale to biological levels".[289]

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Project focused on hardware development related to AGI, not AGI itself

[281] http://humancompatible.ai/about
[282] http://www.chrest.info/history.html
[283] http://www.chrest.info
[284] Lane and Gobet (2012)
[285] http://chrest.info/publications.html
[286] http://chrest.info
[287] http://www.research.ibm.com/articles/brain-chip.shtml
[288] Merolla et al. (2014)
[289] http://www.darpa.mil/program/systems-of-neuromorphic-adaptive-plastic-scalable-electronics

Cognitive Systems Toolkit (CST)

Main website: http://cst.fee.unicamp.br

CST is a project led by Ricardo Gudwin of University of Campinas in Brazil. It is "a Java-based toolkit to allow the construction of Cognitive Architectures".[290]

Reason for consideration: A project related to technical aspects of AGI

Reason for exclusion: Project focused on tools that could be used to develop AGI, not on building an AGI

Comirit

Main website: http://www.comirit.com

Comirit was a project of Benjamin Johnston of University of Technology Sydney. It aimed to build "robotic and software systems with commonsense intelligence" with a short-term focus of "weeks or months, rather than decades", but is "inspired by the long term goals of creating systems that have deep human-like understanding of the real world", and thus pursued designs that "can be gradually evolved into more capable systems".[291] It was active mainly around 2010-2011.

Reason for consideration: An R&D project with AGI aspirations

Reason for exclusion: Apparently inactive

Dav & SAIL

Main website: http://www.cse.msu.edu/~weng/research/LM.html

Dav & SAIL were projects of Juyang Weng of Michigan State University. The two robots were designed to learn as human children do. The project was aimed at achieving "machine's human-level performance through autonomous development".[292] The work received funding from the National Science Foundation and DARPA.[293]

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Apparently inactive

DUAL

Main website: http://alexpetrov.com/proj/dual

[290] http://cst.fee.unicamp.br
[291] http://www.comirit.com
[292] http://www.cse.msu.edu/~weng/research/LM.html
[293] Weng et al. (1999)

DUAL was led by Boicho Kokinov at New Bulgarian University. It was active from around 1999 to 2005. It was based on Marvin Minsky's *Society of Mind*, in which minds are made from interacting sub-mind "agents". DUAL integrates symbolic and emergentist approaches and also integrates declarative learning (learning about information one can readily talk about) and procedural learning (learning that is more habitual and harder to talk about).[294]

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Apparently inactive

Entropica

Main website: http://entropica.com

Entropica is an AGI project and private company led by Alexander Wissner-Gross. It appears to have been founded in 2013.[295] It was included in a 2017 list of AGI projects.[296] However, no activity since 2013 was found, and in August 2017 Wissner-Gross did not include Entropica on a list of companies on his website[297] or on his CV.[298] Entropica is based on ideas Wissner-Gross published in a research paper in the same year.[299] A video describing the project states that it is "broadly applicable to a variety of domains" and shows it functioning in several seemingly different domains.[300] The video is the only content on the project's website. Media coverage described it as a breakthrough to AGI and superintelligence,[301] but other AI researchers have been critical[302] and some observers suspect it to be a hoax.[303]

Reason for consideration: An AGI R&D project

Reason for exclusion: Apparently inactive

Einstein

Main website: https://www.salesforce.com/products/einstein

Einstein is a project of SalesForce that applies AI to its customer service business. Einstein grew out of the private company MetaMind, which SalesForce acquired in 2016.[304] Chief Scientist Richard Socher reportedly aspires to build AGI.[305] However, no indications were found that Einstein is working on AGI.

[294] http://alexpetrov.com/proj/dual
[295] https://www.ventureradar.com/organisation/Entropica/ff1cc5d7-f82a-4a9c-b951-cfd9747ff310
[296] http://2ai.org/landscape
[297] http://www.alexwg.org/companies
[298] http://www.alexwg.org/AWG-CV.pdf
[299] Wissner-Gross and Freer (2013)
[300] https://www.youtube.com/watch?v=cT8ZqChv8P0
[301] Dvorsky (2013)
[302] Marcus and Davis (2013)
[303] https://www.quora.com/How-can-we-prove-that-Entropica-is-a-hoax
[304] Novet (2016)
[305] The Economist (2016)

Reason for consideration: An AI R&D project led by someone who seeks to build AGI

Reason for exclusion: No AGI R&D found

EPIC

Main website: None

EPIC was led by computer scientist David Kieras at University of Michigan. Goertzel (2014, p.16) writes that "It has been connected to SOAR for problem solving, planning and learning."

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Apparently inactive

EvoGrid

Main website: http://www.evogrid.org

EvoGrid was an open source artificial life project initiated by Bruce Damer. It sought to overcome the computing challenge of artificial life by accessing a distributed network of computer hardware similar to that used by projects like SETI@home. The project website shows no updates since around 2010, though Damer's work on artificial life continues.[306]

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Apparently inactive

GLAIR

Main website: none found

GLAIR was a project of Stuart Shapiro of State University of New York at Buffalo. It was active from around 1993-2013.[307] GLAIR aimed for "computational understanding and implementation of human-level intelligent behavior".[308]

Reason for consideration: An AGI R&D project

Reason for exclusion: Apparently inactive

GMU BICA

[306] http://www.damer.com
[307] https://www.cse.buffalo.edu/~shapiro/Papers
[308] Shapiro and Bona (2010, p. 307)

Main website: http://mason.gmu.edu/~asamsono/bica.html

GMU Bica was a project of Alexei Samsonovich of George Mason University. It was active around 2006-2007. Its website describes it as "a general cognitive architecture" based on self-awareness.[309]

Reason for consideration: An AGI R&D project

Reason for exclusion: Apparently inactive

Goedel Machine

Main website: http://people.idsia.ch/~juergen/goedelmachine.html

Goedel Machine was a project of Jürgen Schmidhuber of the Dalle Molle Institute for Artificial Intelligence Research in Switzerland. The Goedel Machine proceeds by taking the action it proves to be best at each step in its activity, which requires infinite computing power (Goertzel 2014, p.25). Schmidhuber writes on his website that "Since age 15 or so, the main goal of professor Jürgen Schmidhuber has been to build a self-improving Artificial Intelligence (AI) smarter than himself, then retire".[310]

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Apparently inactive; project focused on theoretical aspects of AGI, not the development of an AGI

Google Brain

Main website: https://research.google.com/teams/brain

Google Brain is an AI research group at Google. Its researchers have collaborated with DeepMind on AGI research,[311] but its work is mainly focused on deep learning.[312]

Reason for consideration: A research group with links to AGI

Reason for exclusion: Not sufficiently focused on AGI

HUMANOBS

Main website: http://www.humanobs.org

HUMANOBS was a project for developing robots that "can learn social interaction", which it states is "a big step towards the ultimate goal of creating intelligence that is both self-sufficient and adaptable in a wide variety of environments." It was active from 2009 to 2014 and participated in several AGI

[309] http://mason.gmu.edu/~asamsono/bica.html
[310] http://people.idsia.ch/~juergen
[311] Fernando et al. (2017)
[312] https://research.google.com/teams/brain/about.html

conferences. It was run by a consortium of universities and research institutes in Iceland, Italy, Spain, Switzerland, and the United Kingdom.[313]

Reason for consideration: A former AGI R&D project

Reason for exclusion: Apparently inactive

IM-CLEVER

Main website: http://www.im-clever.eu

IM-CLEVER was a project to design robots that could learn on their own and apply their knowledge across contexts. It has been inactive since around 2013. It was led by Gianluca Baldassarre of Istituto di Scienze e Tecnologie della Cognizione (Institute of Cognitive Sciences and Technologies) in Italy and includes collaborators from around Europe and one group in the United States.[314] It worked on a robotics platform "iCub" for robots that could learn new skills and apply them to diverse tasks.[315]

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Apparently inactive

Intel

Main website: https://intel.com (no dedicated AI website found)

Intel is active in AI research.[316] It has acquired several AI companies.[317] However, there is no apparent focus on AGI.

Reason for consideration: Intel is a major computing technology company

Reason for exclusion: No indications of AGI projects were found

Israel Brain Technologies (IBT)

Main website: http://israelbrain.org

IBT is a neuroscience research project. It is an Israeli nonprofit based in Ramat HaSharon. It receives funding from the Israeli government, philanthropists, and corporations.[318] Its mission is "to accelerate the development of innovative treatments and cures for brain disease".[319]

[313] http://www.humanobs.org
[314] http://www.im-clever.eu/homepage/project/partners/partners
[315] http://www.im-clever.eu
[316] https://www.intel.com/content/www/us/en/analytics/artificial-intelligence/overview.html
[317] Tamturk (2017)
[318] http://israelbrain.org/donate
[319] http://israelbrain.org/about-us/mission

Reason for consideration: A large-scale brain research project, similar to Blue Brain

Reason for exclusion: Project focused on basic neuroscience, not the development of an AGI

Kuipers Group

Main website: https://www.cs.utexas.edu/users/qr/robotics/bootstrap-learning.html

Benjamin Kuipers of the University of Texas led a group that developed robots that would learn a wide range of information about the world on their own from their own experiences. They published from 1997-2007.

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Apparently inactive

Large-Scale Model of Mammalian Thalamocortical Systems (LSMMTS)

Main website: https://www.izhikevich.org/publications/large-scale_model_of_human_brain.htm

LSMMTS was a project of Eugene Izhikevich and Gerald Edelman of The Neurosciences Institute. LSMMTS is notable for being a brain simulation "on a scale similar to that of the full human brain itself" (Goertzel 2014, p.19).

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Apparently inactive

Machine Intelligence Research Institute (MIRI)

Main website: https://intelligence.org

MIRI is an independent nonprofit research group focused on "foundational mathematical research to ensure smarter-than-human artificial intelligence has a positive impact."[320] It states that its mission is "to develop formal tools for the clean design and analysis of general-purpose AI systems, with the intent of making such systems safer and more reliable when they are developed."[321]

Reason for consideration: A research project focused on technical details of AGI

Reason for exclusion: Project focused safety aspects of AGI, not on the development of an AGI

Neurogrid

Main website: https://web.stanford.edu/group/brainsinsilicon/neurogrid.html

[320] https://intelligence.org
[321] https://intelligence.org/about

Neurogrid is computer hardware designed for running low-cost brain simulations. It is led by Kwabena Boahen of Stanford University's Bioengineering department. A 2014 version of Neurogrid claims to be "9,000 times faster and using significantly less power than a typical PC" but still much less energy-efficient than the human brain.[322]

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Project focused on hardware development related to AGI, not AGI itself

NOMAD (Neurally Organized Mobile Adaptive Device)

Main website: http://www.nsi.edu/~nomad

NOMAD was a project of The Neurosciences Institute, a nonprofit research institute in La Jolla, California led by Nobel Laureate Gerald Edelman. NOMAD used "large numbers of simulated neurons evolving by natural selection" to achieve various tasks Goertzel (2014, p.17).

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Apparently inactive

OSCAR

Main website: http://johnpollock.us/ftp/OSCAR-web-page/oscar.html

OSCAR was a project of John Pollock of the University of Arizona, active from around 1995 to 2005. (Pollock passed away in 2009.[323]) Pollock writes, "The 'grand problem' of AI has always been to build artificial agents of human-like intelligence… capable of operating in environments of real-world complexity… OSCAR is a cognitive architecture for GIAs [generally intelligent agents], implemented in LISP."[324]

Reason for consideration: An AGI R&D project

Reason for exclusion: Apparently inactive

PAGI World

Main website: http://rair.cogsci.rpi.edu/projects/pagi-world

[322] http://news.stanford.edu/pr/2014/pr-neurogrid-boahen-engineering-042814.html, a press release discussing Benjamin et al. (2014)
[323] http://johnpollock.us
[324] Pollock (2008)

PAGI World is a project led by John Licato of University of South Florida and based at Rensselaer Polytechnic Institute, where Licato was a Ph.D. student. PAGI world is "a simulation environment written in Unity 2D which allows AI and AGI researchers to test out their ideas".[325]

Reason for consideration: A project on an aspect of AGI R&D

Reason for exclusion: Project focused on tools for evaluating AGI, not on developing an AGI

PolyScheme

Main website: https://dspace.mit.edu/handle/1721.1/8325

PolyScheme was developed for the Ph.D. thesis of Nicholas Cassimatis at Massachusetts Institute of Technology. It "integrates multiple methods of representation, reasoning and inference schemes for general problem solving" (Goertzel 2014, p.24). The project is no longer active.

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Apparently inactive

Quantum Artificial Intelligence Lab (QAIL)

Main website: https://research.googleblog.com/2013/05/launching-quantum-artificial.html

QAIL is a project at Google seeking to use quantum computing to advance AI. QAIL is listed as an AGI project by 2AI,[326] and some commentators propose that quantum computing could be important for developing AGI.[327] However, QAIL gives no indications of aiming for AGI.

Reason for consideration: Listed as an AGI project by 2AI

Reason for exclusion: No apparent AGI focus

SAL (Synthesis of ACT-R and Leabra)

Main website: None

SAL was developed by David Jilk, Christian Lebiere and colleagues at the eCortex corporation of Boulder, Colorado, the University of Colorado and Carnegie Mellon University. The project produced a brief publishing record[328] and appears inactive since 2008. As its name implies, SAL is based on ACT-R (see dedicated ACT-R entry) and Leabra, a neural simulation.

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

[325] http://rair.cogsci.rpi.edu/projects/pagi-world
[326] http://2ai.org/landscape
[327] For example, Wang (2014), DeAngelis (2014)
[328] E.g., Jilk et al. (2008)

Reason for exclusion: Apparently inactive

Scene Based Reasoning (SBR)

Main website: http://agi-conf.org/2015/wp-content/uploads/2015/07/agi15_bergmann.pdf

SBR is an AGI R&D project by Frank Bergmann and Brian Fenton presented at the 2015 AGI conference, but apparently inactive since.

Reason for consideration: An AGI R&D project

Reason for exclusion: Apparently inactive

Shruti

Main website: http://www1.icsi.berkeley.edu/~shastri/shruti

Shruti was a project of Lokendra Shastri of University of California, Berkeley. It was active from around 1996-2007. The project developed computational tools based on fast, reflex-like human inference. It was been funded by, among others, the US National Science Foundation, Office of Naval Research, and Army Research Institute.[329]

Reason for consideration: Listed in the AGI review paper Goertzel (2014)

Reason for exclusion: Apparently inactive

Twitter

Main website: https://cortex.twitter.com

Twitter has an AI group called Cortex but no indications of AGI work was found.

Reason for consideration: Twitter is a major computing technology company

Reason for exclusion: No indications of AGI projects were found

Xapagy

Main website: http://www.xapagy.com

Xapagy is a cognitive architecture designed "to perform narrative reasoning, that is to model / mimic the mental processes humans perform with respect to stories".[330]

[329] http://www1.icsi.berkeley.edu/~shastri/shruti
[330] http://www.xapagy.com/?page_id=26

Reason for consideration: A cognitive architecture presented at an AGI conference[331]

Reason for exclusion: The focus on narrative makes it ultimately narrow AI

Ymir

Main website: http://alumni.media.mit.edu/~kris/ymir.html

Ymir was a project of Kristinn Thórisson. It was active from around 1999-2009.[332] It "was created with the goal of endowing artificial agents with human-like communicative and manipulation capabilities in the form of embodied multimodal task-oriented dialog skills".[333]

Reason for consideration: An AGI R&D project

Reason for exclusion: Apparently inactive

[331] Bölöni (2014)
[332] http://alumni.media.mit.edu/~kris/ymir.html
[333] Thórisson and Helgasson (2012), p.8

References

Allen G, Kania EB, 2017. China is using America's own plan to dominate the future of artificial intelligence. Foreign Policy, 8 September, http://foreignpolicy.com/2017/09/08/china-is-using-americas-own-plan-to-dominate-the-future-of-artificial-intelligence

Amodei D, Olah C, Steinhardt J, Christiano P, Schulman J, Mané D, 2016. Concrete problems in AI safety. arXiv:1606.06565.

Architecht, no date. Microsoft has an AI coming-out party. Architecht, Issue #111, http://news.architecht.io/issues/microsoft-has-an-ai-coming-out-party-65494

Armstrong S, Bostrom N, Shulman C, 2016. Racing to the precipice: A model of artificial intelligence development. AI & Society, 31(2), 201-206.

Arrabales R, Muñoz J, 2010. The awakening of conscious bots: Inside the mind of the 2K BotPrize 2010 winner. AiGameDev, 21 October, https://aigamedev.com/open/articles/conscious-bot

Arrabales R, Muñoz J, Ledezma A, Gutierrez G, Sanchis A, 2013. A machine consciousness approach to the design of human-like bots. In Hingston P (Ed), Believable Bots. Berlin: Springer, pp. 171-191.

Auerbach C, Silverstein LB, 2003. Qualitative Data: An Introduction to Coding and Analysis. New York: NYU Press.

Bach J, 2015. Modeling motivation in MicroPsi 2. In Bieger J, Goertzel B, Potapov A (Eds), Proceedings of AGI 2015, 8th International Conference on Artificial General Intelligence. Cham, Switzerland: Springer, pp. 3-13.

Barrett AM, Baum SD, 2017a. A model of pathways to artificial superintelligence catastrophe for risk and decision analysis. Journal of Experimental & Theoretical Artificial Intelligence, 29(2), 397-414.

Barrett AM, Baum SD, 2017b. Risk analysis and risk management for the artificial superintelligence research and development process. In Callaghan V, Miller J, Yampolskiy R, Armstrong S (Eds), The Technological Singularity. Berlin: Springer, pp. 127-140.

Baum SD, 2017a. Reconciliation between factions focused on near-term and long-term artificial intelligence. AI & Society, in press, doi 10.1007/s00146-017-0734-3.

Baum SD, 2017b. Social choice ethics in artificial intelligence. AI & Society, in press, doi 10.1007/s00146-017-0760-1.

Baum SD, 2017c. On the promotion of safe and socially beneficial artificial intelligence. AI & Society, 32(4), 543-551.

Baum SD, Goertzel B, Goertzel TG, 2011. How long until human-level AI? Results from an expert assessment. Technological Forecasting and Social Change 78(1), 185-195.

Benjamin BV, Gao P, McQuinn E, Choudhary S, Chandrasekaran AR, Bussat JM, et al., 2014. Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations. Proceedings of the IEEE, 102(5), 699-716.

Best J, 2016. Startup Vicarious is aiming to build the first general artificial intelligence system—just don't expect it any time soon. ZDNet, 2 March, http://www.zdnet.com/article/the-ghost-in-the-machine-vicarious-and-the-search-for-ai-that-can-rival-the-human-brain

Bieger J, Thórisson KR, Wang P, 2015. Safe baby AGI. In Bieger J, Goertzel B, Potapov A (Eds), Proceedings of AGI 2015, 8th International Conference on Artificial General Intelligence. Cham, Switzerland: Springer, pp. 46-49.

Bölöni L, 2014. Autobiography based prediction in a situated AGI agent. In Goertzel B, Orseau L, Snaider J (Eds), Proceedings of AGI 2014, 7th International Conference on Artificial General Intelligence. Cham, Switzerland: Springer, pp. 11-20.

Bonhoff GM, 2008. Using hierarchical temporal memory for detecting anomalous network activity. Masters Thesis, Department of Electrical and Computer Engineering, US Air Force Institute of Technology.
Boyle A, 2016. A bot is born: Kimera Systems adds 'Nigel' to the crowd of AI assistants. GeekWire, 9 August, https://www.geekwire.com/2016/kimera-systems-nigel-ai-agi
Brandstätter C, Dietrich D, Doblhammer K, Fittner M, Fodor G, Gelbard F, et al., 2015. Natural Scientific, Psychoanalytical Model of the Psyche for Simulation and Emulation. Scientific Report III, Institute of Computer Technology, Vienna University of Technology, Version eV005.
Bringsjord S, 2012. Belief in the singularity is logically brittle. Journal of Consciousness Studies, 19(7), 14-20.
Bostrom N, 2014. Superintelligence: Paths, Dangers, Strategies. Oxford: Oxford University Press.
Bostrom N, 2017. Strategic implications of openness in AI development. Global Policy, 8(2), 135-148.
Chamberlain L, 2016. Uber invests heavily in artificial intelligence with launch of Uber AI Labs. GeoMarketing, 5 December, http://www.geomarketing.com/uber-invests-in-artificial-intelligence-uber-ai-labs
Chen X, 2017. Gary Marcus: The road to artificial general intelligence. Medium, 1 July, https://medium.com/@Synced/gary-marcus-the-road-to-artificial-general-intelligence-ce5b1371aa02
Choi D, Langley P, 2017. Evolution of the Icarus cognitive architecture. Cognitive Systems Research, in press, doi 10.1016/j.cogsys.2017.05.005.
Cutler K-M, 2014. Vicarious grabs a huge, new $40M growth round to advance artificial intelligence. TechCrunch, https://techcrunch.com/2014/03/22/vicarious-grabs-a-huge-new-40m-growth-round-to-advance-artificial-intelligence
Chong HQ, Tan AH, Ng GW, 2007. Integrated cognitive architectures: A survey. Artificial Intelligence Review, 28(2), 103.
Christiano P, Leike J, Brown TB, Martic M, Legg S, Amodei D, 2017. Deep reinforcement learning from human preferences. arXiv:1706.03741.
Coffey A, Atkinson P, 1996. Making Sense of Qualitative Data: Complementary Research Strategies. Thousand Oaks, CA: Sage.
Conn A, 2016. The White House considers the future of AI. Future of Life Institute, 15 June, https://futureoflife.org/2016/06/15/white-house-future-ai-1
de Garis H, 2004. The Artilect War. Pittsburgh: Etc Press.
de Garis H, Shuo C, Goertzel B, Ruiting L, 2010. A world survey of artificial brain projects, Part I: Large-scale brain simulations. Neurocomputing, 74(1), 3-29.
DeAngelis SF, 2014. Quantum computers will transform the world—we hope. Enterra Insights Blog, 24 December, http://www.enterrasolutions.com/quantum-computers-will-transform-world-hope
Deaton C, Shepard B, Klein C, Mayans C, Summers B, Brusseau A, et al., 2005. The comprehensive terrorism knowledge base in Cyc. In Proceedings of the 2005 International Conference on Intelligence Analysis.
Dewey D, 2015. Long-term strategies for ending existential risk from fast takeoff. In Müller VC (Ed), Risks of Artificial Intelligence. Boca Raton: CRC, pp. 243-266.
Dong D, Franklin S, 2014. The action execution process implemented in different cognitive architectures: A review. Journal of Artificial General Intelligence, 5(1), 49-68.
Duch W, Oentaryo RJ, Pasquier M, 2008. Cognitive architectures: Where do we go from here? In Wang P, Goertzel B, Franklin S (Eds), Frontiers in Artificial Intelligence and Applications, Vol. 171. Amsterdam: IOS Press, pp. 122-136.
Dvorsky G, 2013. How Skynet might emerge from simple physics. io9, 26 April, http://io9.gizmodo.com/how-skynet-might-emerge-from-simple-physics-482402911

Etherington D, 2017. Microsoft creates an AI research lab to challenge Google and DeepMind. TechCrunch, 12 July, https://techcrunch.com/2017/07/12/microsoft-creates-an-ai-research-lab-to-challenge-google-and-deepmind

Etzioni O, 2016. No, the experts don't think superintelligent AI is a threat to humanity. MIT Technology Review, 20 September, https://www.technologyreview.com/s/602410/no-the-experts-dont-think-superintelligent-ai-is-a-threat-to-humanity

Fernando C, Banarse D, Blundell C, Zwols Y, Ha D, Rusu AA, et al., 2017. PathNet: Evolution channels gradient descent in super neural networks. arXiv:1701.08734.

Gershgorn D, 2017. China is funding Baidu to take on the US in deep-learning research. Quartz, 22 February, https://qz.com/916738/china-is-funding-baidu-to-take-on-the-united-states-in-deep-learning-research

Gibbs S, 2014. Google buys UK artificial intelligence startup Deepmind for £400m. The Guardian, 27 January, https://www.theguardian.com/technology/2014/jan/27/google-acquires-uk-artificial-intelligence-startup-deepmind

Goertzel B, 2008. Ben Goertzel reports from Xiamen China. Institute for Ethics and Emerging Technologies, June 20, https://ieet.org/index.php/IEET2/more/goertzel20080620

Goertzel B, 2009. Will China build AGI first? The Multiverse According to Ben, 26 July, http://multiverseaccordingtoben.blogspot.com/2009/07/will-china-build-agi-first.html

Goertzel B, 2010. A Cosmist Manifesto: Practical Philosophy for the Posthuman Age. Humanity+ Press.

Goertzel B, 2011. Pei Wang on the path to artificial general intelligence. h+ Magazine, 27 January, http://hplusmagazine.com/2011/01/27/pei-wang-path-artificial-general-intelligence

Goertzel B, 2012a. Should humanity build a global AI nanny to delay the singularity until it's better understood? Journal of Consciousness Studies, 19(1-2), 96-111.

Goertzel B, 2012b. AI and Ethiopia: An unexpected synergy. KurzweilAI, 25 October, http://www.kurzweilai.net/ai-and-ethiopia-an-unexpected-synergy

Goertzel B, 2014. Artificial general intelligence: Concept, state of the art, and future prospects. Journal of Artificial General Intelligence, 5(1), 1-48.

Goertzel B, 2015. Superintelligence: Fears, promises and potentials. Journal of Evolution and Technology, 25(2), 55-87.

Goertzel B, 2016. Infusing advanced AGIs with human-like value systems: Two theses. Journal of Evolution and Technology, 26(1), 50-72.

Goertzel B, 2017a. The corporatization of AI is a major threat to humanity. H+ Magazine, July 21, http://hplusmagazine.com/2017/07/21/corporatization-ai-major-threat-humanity

Goertzel B, 2017b. SingularityNET—AGI of the People, by the People and for the People (and the Robots!). Medium, 26 October, https://medium.com/ben-goertzel-on-singularitynet/singularitynet-agi-of-the-people-by-the-people-and-for-the-people-and-the-robots-7246a6886e90

Goertzel B, Pitt J, 2012. Nine ways to bias open-source AGI toward friendliness. Journal of Evolution & Technology, 22(1), 116-131.

Goertzel B, Lian R, Arel I, De Garis H, Chen S, 2010. A world survey of artificial brain projects, Part II: Biologically inspired cognitive architectures. Neurocomputing, 74(1), 30-49.

Goertzel B, Giacomelli S, Hanson D, Pennachin C, Argentieri M, the SingularityNET team, 2017. SingularityNET: A decentralized, open market and inter-network for AIs. SingularityNET, 1 November, https://public.singularitynet.io/whitepaper.pdf

Good IJ, 1965. Speculations concerning the first ultraintelligent machine. Advances in Computers, 6, 31-88.

Griffin M, 2017. Baidu's AI just achieved zero shot learning. Fanatical Futurist, 14 April, http://www.fanaticalfuturist.com/2017/04/baidus-ai-just-achieved-zero-shot-learning

Hawkins J, 2015. The Terminator is not coming. The future will thank us. recode, 2 March, https://www.recode.net/2015/3/2/11559576/the-terminator-is-not-coming-the-future-will-thank-us
Hawkins J, 2017. What intelligent machines need to learn from the neocortex. IEEE Spectrum, 2 June, http://spectrum.ieee.org/computing/software/what-intelligent-machines-need-to-learn-from-the-neocortex
Heath A, 2017. Mark Zuckerberg's plan to create non-voting Facebook shares is going to trial in September. Business Insider, May 4, http://www.businessinsider.com/trial-challenge-facebook-non-voting-shares-set-september-2017-2017-5
Hellman D, Hellman M, 2016. A New Map For Relationships: Creating True Love At Home and Peace On The Planet. New Map Publishing.
Hibbard B, 2002. Super-Intelligent Machines. New York: Springer.
High P, 2016. Vicarious is the AI company that includes Zuckerberg, Bezos, Musk, and Thiel as investors. Forbes, 11 April, https://www.forbes.com/sites/peterhigh/2016/04/11/vicarious-is-the-ai-company-that-includes-zuckerberg-bezos-musk-and-thiel-as-investors
Hughes JJ, 2007. Global technology regulation and potentially apocalyptic technological threats. In Allhoff F (Ed), Nanoethics: The Ethical and Social Implications of Nanotechnology. Hoboken, NJ: John Wiley, pp. 201-214.
Ingram M, 2017. At Alphabet, there are only two shareholders who matter. Fortune, June 7, http://fortune.com/2017/06/07/alphabet-shareholders-meeting
Jee C, 2016. What is artificial general intelligence? And does Kimera Systems' 'Nigel' qualify? TechWorld, 26 August, https://www.techworld.com/data/what-is-artificial-general-intelligence-3645268
Jilk DJ, Lebiere C, O'Reilly RC, Anderson JR, 2008. SAL: An explicitly pluralistic cognitive architecture. Journal of Experimental and Theoretical Artificial Intelligence, 20(3), 197-218.
Joy B, 2000. Why the future doesn't need us. Wired, 8(4), 238-263.
Kaczynski T, 1995. Industrial Society And Its Future.
Kotseruba I, Gonzalez OJA, Tsotsos JK, 2016. A review of 40 years of cognitive architecture research: Focus on perception, attention, learning and applications. arXiv:1610.08602.
Lane PC, Gobet F, 2012. CHREST models of implicit learning and board game interpretation. In Bach J, Goertzel B, Ikle M (Eds), Proceedings of AGI 2012, 5th International Conference on Artificial General Intelligence. Berlin: Springer, pp. 148-157.
Langley P, 2017. Progress and challenges in research on cognitive architectures. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), pp. 4870-4876.
Langley P, Laird J, Rogers S, 2008. Cognitive architectures: Research issues and challenges. Cognitive Systems Research 10, 141-160.
LeCun Y, Bengio Y, Hinton G, 2015. Deep learning. Nature, 521(7553), 436-444.
Lee HS, Betts S, Anderson JR, 2017. Embellishing problem-solving examples with deep structure information facilitates transfer. Journal of Experimental Education, 85(2), 309-333.
Lentzos F, 2011. Hard to prove: The verification quandary of the Biological Weapons Convention. Nonproliferation Review, 18(3), 571-582.
Lipsitch M, Inglesby TV, 2014. Moratorium on research intended to create novel potential pandemic pathogens. mBio, 5(6), e02366-14.
Love D, 2014. The most ambitious artificial intelligence project in the world has been operating in near secrecy for 30 years. Business Insider, 2 July, http://www.businessinsider.com/cycorp-ai-2014-7
Maddox T, 2016. Should Baidu be your AI and machine learning platform? ZDNet, 1 December, http://www.zdnet.com/article/should-baidu-be-your-ai-and-machine-learning-platform
Madl T, Franklin S, 2015. Constrained incrementalist moral decision making for a biologically inspired cognitive architecture. In Trappl R (Ed), A Construction Manual for Robots' Ethical Systems: Requirements, Methods, Implementations. Cham, Switzerland: Springer, pp. 137-153.

Mannes J, 2017. Tencent to open AI research center in Seattle. TechCrunch, 28 April, https://techcrunch.com/2017/04/28/tencent-to-open-ai-research-center-in-seattle
Marcus G, 2017. Artificial intelligence is stuck. Here's how to move it forward. New York Times, July 29. https://nytimes.com/2017/07/29/opinion/sunday/artificial-intelligence-is-stuck-heres-how-to-move-it-forward.html
Marcus G, Davis E, 2013. A grand unified theory of everything. New Yorker, 6 May, http://www.newyorker.com/tech/elements/a-grand-unified-theory-of-everything
Marquis C, Toffel, MW, Zhou Y, 2016. Scrutiny, norms, and selective disclosure: A global study of greenwashing. Organization Science, 27(2), 483-504.
McDermott D, 2012. Response to the singularity by David Chalmers. Journal of Consciousness Studies, 19(1-2), 167-172.
Menager D, Choi D, 2016. A robust implementation of episodic memory for a cognitive architecture. In Papafragou A, Grodner D, Mirman D, Trueswell JC (Eds), Proceedings of the 38th Annual Conference of the Cognitive Science Society. Austin: Cognitive Science Society, pp. 620-625.
Merolla PA, Arthur JV, Alvarez-Icaza R, Cassidy AS, Sawada J, Akopyan F, et al., 2014. A million spiking-neuron integrated circuit with a scalable communication network and interface. Science, 345(6197), 668-673.
Metz C, 2016. Uber Buys a mysterious startup to make itself an AI company. Wired, 5 December, https://www.wired.com/2016/12/uber-buys-mysterious-startup-make-ai-company
Muehlhauser L, Helm L, 2012. Intelligence explosion and machine ethics. In Eden A, Søraker J, Moor JH, Steinhart E (Eds), Singularity hypotheses: A Scientific and Philosophical Assessment. Berlin: Springer, pp.101-126.
Nadella S, 2016. The partnership of the future. Slate, 28 June, http://www.slate.com/articles/technology/future_tense/2016/06/microsoft_ceo_satya_nadella_humans_and_a_i_can_work_together_to_solve_society.html
Ng KH, Du Z, Ng GW, 2017. DSO cognitive architecture: Unified reasoning with integrative memory using global workspace theory. In Everitt T, Goertzel B, Potapov A (Eds), Proceedings of AGI 2017, 10th International Conference on Artificial General Intelligence. Cham, Switzerland: Springer, pp. 44-53.
Nivel E, Thórisson KR, Steunebrink BR, Dindo H, Pezzulo G, Rodriguez M, et al., 2013. Bounded recursive self-improvement. Reykjavik University School of Computer Science Technical Report RUTR-SCS13006.
Novet J, 2016. Salesforce forms research group, launches Einstein A.I. platform that works with Sales Cloud, Marketing Cloud. Venture Beat, 18 September, https://venturebeat.com/2016/09/18/salesforce-forms-research-group-launches-einstein-a-i-platform-that-works-with-sales-cloud-marketing-cloud
O'Reilly RC, Hazy TE, Herd SA, 2016. The Leabra cognitive architecture: How to play 20 principles with nature. In Chipman SEF (Ed), The Oxford Handbook of Cognitive Science. Oxford: Oxford University Press, pp.91-116.
Oreskes N, Conway EM, 2010. Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming. New York: Bloomsbury.
Oudeyer PY, Ly O, Rouanet P, 2011. Exploring robust, intuitive and emergent physical human-robot interaction with the humanoid Acroban. IEEE-RAS International Conference on Humanoid Robots, Bled, Slovenia.
Pollock JL, 2008. OSCAR: An agent architecture based on defeasible reasoning. Proceedings of the 2008 AAAI Spring Symposium on Architectures for Intelligent Theory-Based Agents, pp. 55-60.
Poo MM, Du JL, Ip NY, Xiong ZQ, Xu B, Tan T, 2016. China Brain Project: Basic neuroscience, brain diseases, and brain-inspired computing. Neuron, 92(3), 591-596.
Posner EA, 2014, The Twilight of Human Rights Law. Oxford: Oxford University Press.

Potapov A, Rodionov S, 2014. Universal empathy and ethical bias for artificial general intelligence. Journal of Experimental & Theoretical Artificial Intelligence, 26(3), 405-416.
Potapov A, Rodionov S, Potapova V, 2016. Real-time GA-based probabilistic programming in application to robot control. In Steunebrink B, Wang P, Goertzel B (Eds), Proceedings of AGI 2016, 9th International Conference on Artificial General Intelligence. Cham, Switzerland: Springer, pp. 95-105.
Real AI, 2017. Prosaic AI. http://realai.org/prosaic
Rosenbloom PS, 2013. The Sigma cognitive architecture and system. AISB Quarterly, 136, 4-13.
RT, 2017. 'Whoever leads in AI will rule the world': Putin to Russian children on Knowledge Day. RT, 1 September, https://www.rt.com/news/401731-ai-rule-world-putin
Samsonovich AV, 2010. Toward a unified catalog of implemented cognitive architectures. Biologically Inspired Cognitive Architectures, 221, 195-244.
Scherer MU, 2016. Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. Harvard Journal of Law & Technology, 29(2), 353-400
Schienke EW, Tuana N, Brown DA, Davis KJ, Keller K, Shortle JS, et al., 2009. The role of the NSF Broader Impacts Criterion in enhancing research ethics pedagogy. Social Epistemology, 23(3-4), 317-336.
Shapiro SC, Bona JP, 2010. The GLAIR cognitive architecture. International Journal of Machine Consciousness, 2(2), 307-332.
Shapiro SC, Schlegel DR, 2016. Natural language understanding for information fusion. In Rogova G, Scott P (Eds), Fusion Methodologies in Crisis Management. Cham, Switzerland: Springer, pp. 27-45.
Shead S, 2017. DeepMind's army of AI researchers in Mountain View is now over 20 people strong. Business Insider, 31 May, http://www.businessinsider.com/deepminds-small-army-of-ai-researchers-in-mountain-view-is-growing-2017-5
Shulman C, 2009. Arms control and intelligence explosions. In 7th European Conference on Computing and Philosophy (ECAP), Bellaterra, Spain, July 2-4.
Skaba W, 2012a. Binary space partitioning as intrinsic reward. In Bach J, Goertzel B, Ikle M (Eds), Proceedings of AGI 2012, 5th International Conference on Artificial General Intelligence. Berlin: Springer, pp. 282-291.
Skaba W, 2012b. The AGINAO self-programming engine. Journal of Artificial General Intelligence, 3(3), 74-100.
Sotala K, Yampolskiy RV, 2014. Responses to catastrophic AGI risk: a survey. Physica Scripta, 90(1), 018001.
Starzyk JA, Graham J, 2015. MLECOG: Motivated learning embodied cognitive architecture. IEEE Systems Journal, 11(3), 1272-1283.
Steunebrink BR, Thórisson KR, Schmidhuber J, 2016. Growing recursive self-improvers. In Steunebrink B, Wang P, Goertzel B (Eds), Proceedings of AGI 2016, 9th International Conference on Artificial General Intelligence. Cham, Switzerland: Springer, pp. 129-139.
Stilgoe J, Maynard A, 2017. It's time for some messy, democratic discussions about the future of AI. The Guardian, 1 February, https://www.theguardian.com/science/political-science/2017/feb/01/ai-artificial-intelligence-its-time-for-some-messy-democratic-discussions-about-the-future
Strannegård C, Nizamani AR, 2016. Integrating symbolic and sub-symbolic reasoning. In Steunebrink B, Wang P, Goertzel B (Eds), Proceedings of AGI 2016, 9th International Conference on Artificial General Intelligence. Cham, Switzerland: Springer, pp. 171-180.
Strannegård C, Nizamani AR, Juel J, Persson U, 2016. Learning and reasoning in unknown domains. Journal of Artificial General Intelligence, 7(1), 104-127.

Strannegård C, Svangård N, Bach J, Steunebrink B, 2017a. Generic animats. In Everitt T, Goertzel B, Potapov A (Eds), Proceedings of AGI 2017, 10th International Conference on Artificial General Intelligence. Cham, Switzerland: Springer, pp. 23-32.
Strannegård C, Svangård N, Lindström D, Bach J, Steunebrink B, 2017b. The animat path to artificial general intelligence. In Proceedings of IJCAI-17 Workshop on Architectures for Generality & Autonomy.
Taatgen N, Anderson J, 2010. The past, present, and future of cognitive architectures. Topics in Cognitive Science 2(4), 693-704.
Tamturk V, 2017. Google, Apple, Facebook, and Intel battle for AI supremacy. CMS Connected, 21 April. http://www.cms-connected.com/News-Archive/April-2017/Google-Apple-Facebook-Intel-Microsoft-Salesforce-Twitter-Battle-AI-Supremacy
Temperton J, 2016. Uber acquires AI firm to boost self-driving car project. Wired, 6 December, http://www.wired.co.uk/article/uber-artificial-intelligence-lab-self-driving-cars
The Economist, 2016. From not working to neural networking. The Economist, 25 June, https://www.economist.com/news/special-report/21700756-artificial-intelligence-boom-based-old-idea-modern-twist-not
Theil S, 2015. Why the human brain project went wrong—and how to fix it. Scientific American, 1 October, https://www.scientificamerican.com/article/why-the-human-brain-project-went-wrong-and-how-to-fix-it
Thórisson K, Helgasson H, 2012. Cognitive architectures and autonomy: A comparative review. Journal of Artificial General Intelligence, 3(2), 1-30.
Totschnig W, 2017. The problem of superintelligence: Political, not technological. AI & Society, in press, doi 10.1007/s00146-017-0753-0.
Townsend JC, 2016. This startup is teaching machines to think, reason, and communicate like us. Fast Company, 22 November, https://www.fastcompany.com/3065779/this-startup-is-teaching-machines-to-think-reason-and-communicate-like-us
TWiStartups, 2016. Vicarious co-founder Scott Phoenix on innovating in AI & the race to unlock the human brain to create artificial general intelligence, the last tech humans will invent. Medium, 13 May, https://medium.com/@TWiStartups/vicarious-co-founder-scott-phoenix-on-innovating-in-ai-the-race-to-unlock-the-human-brain-to-9161264b0c09
Vanian J, 2017a. Apple's artificial intelligence guru talks about a sci-fi future. Fortune, 28 March, http://fortune.com/2017/03/28/apple-artificial-intelligence
Vanian J, 2017b. Apple just got more public about its artificial intelligence plans. Fortune, 19 July, http://fortune.com/2017/07/19/apple-artificial-intelligence-research-journal
Veness J, Ng KS, Hutter M, Uther W, Silver D, 2011. A Monte-Carlo AIXI approximation. Journal of Artificial Intelligence Research, 40(1), 95-142.
Wang P, 2012. Motivation management in AGI systems. In Bach J, Goertzel B, Ikle M (Eds.), Proceedings of AGI 2012, 5th International Conference on Artificial General Intelligence. Berlin: Springer, pp. 352-361.
Wallach W, Franklin S, Allen C, 2010. A conceptual and computational model of moral decision making in human and artificial agents. Topics in Cognitive Science, 2(3), 454-485.
Wallach W, Allen C, Franklin S, 2011. Consciousness and ethics: Artificially conscious moral agents. International Journal of Machine Consciousness, 3(1), 177-192.
Wang B, 2014. Quantum computing and new approaches to artificial intelligence could get the resources to achieve real breakthroughs in computing. Next Big Future, 31 March, https://www.nextbigfuture.com/2014/03/quantum-computing-and-new-approaches-to.html
Wang P, Li X, 2016. Different conceptions of learning: Function approximation vs. self-organization. In Steunebrink B, Wang P, Goertzel B (Eds.), Proceedings of AGI 2016, 9th International Conference on Artificial General Intelligence. Cham, Switzerland: Springer, pp. 140-149.

Wang P, Talanov M, Hammer P, 2016. The emotional mechanisms in NARS. In Steunebrink B, Wang P, Goertzel B (Eds.), Proceedings of AGI 2016, 9th International Conference on Artificial General Intelligence. Cham, Switzerland: Springer, pp. 150-159.

Webb A, 2017. AI pioneer wants to build the renaissance machine of the future. Bloomberg, 16 January, https://www.bloomberg.com/news/articles/2017-01-16/ai-pioneer-wants-to-build-the-renaissance-machine-of-the-future

Webster G, Creemers R, Triolo P, Kania E, 2017. China's plan to 'lead' in AI: Purpose, prospects, and problems. New America, 1 August, https://www.newamerica.org/cybersecurity-initiative/blog/chinas-plan-lead-ai-purpose-prospects-and-problems

Weng J, Evans CH, Hwang WS, Lee Y-B, 1999. The developmental approach to artificial intelligence: Concepts, developmental algorithms and experimental results. In Proceedings of NSF Design and Manufacturing Grantees Conference, Long Beach, CA.

Wilson G, 2013. Minimizing global catastrophic and existential risks from emerging technologies through international law. Virginia Environmental Law Journal, 31,307-364.

Wissner-Gross AD, Freer CE, 2013. Causal entropic forces. Physical Review Letters, 110(16), 168702.

Yang W, Liu W, Viña A, Tuanmu MN, He G, Dietz T, Liu J, 2013. Nonlinear effects of group size on collective action and resource outcomes. Proceedings of the National Academy of Sciences, 110(27), 10916-10921.

Yampolskiy RV, 2013. Artificial intelligence safety engineering: Why machine ethics is a wrong approach. In Müller VC (Ed), Philosophy and Theory of Artificial Intelligence. Berlin: Springer, pp. 389-396.

Yampolskiy R, Fox J, 2013. Safety engineering for artificial general intelligence. Topoi, 32(2), 217-226.

Yudkowsky E, 2004. Coherent Extrapolated Volition. San Francisco: The Singularity Institute.

Zhang Q, Walsh MM, Anderson JR, 2016. The effects of probe similarity on retrieval and comparison processes in associative recognition. Journal of Cognitive Neuroscience, 29(2), 352-367.