



Newsletter

of the Icelandic Institute for Intelligent Machines, Reykjavik

Volume Four
Issue One, May 2015

COVER

The CADIA Clause is a copyright clause that anyone can insert into their software license. With general copyright laws as the backdrop, the clause restricts users of software bearing the clause from using it for military purposes and unethical activities. Created by Kris Thórisson in collaboration with Brent Britton, the clause comes with its own icon, which illustrates this issue's cover. You may access both at <http://cadia.ru.is/wiki/public:cadia-clause:cadia-clause-main>

ILLUSTRATIONS



Unlike a spider, neurally hard-wired to construct its web according to nature's design, humans are free to design society in whichever way they choose. Like an architect ensuring that a new blueprint is adhered to during a building's construction, researchers producing new knowledge must ensure the safe and sound application of that knowledge. Depicting architectural constructs of various kinds, the illustrative images in this issue drive this point home. (Photos: K. R. Thórisson.)

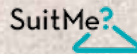
Our industry collaborators include



RÍKISLÖGREGLUSTJÓRINN



Statistics Iceland



Our research partners include



HÁSKÓLI ÍSLANDS



FROM THE DIRECTOR

IS IT ETHICAL TO CREATE INTELLIGENT MACHINES?

BY KRISTINN R. THORISSON

I mentioned in casual conversation to a nuclear physicist colleague of mine that AI is potentially as dangerous as nuclear power. He was mildly offended, of course. The thought stuck with me, only to become relevant much later – a decade later, to be exact. Now, with the likes of Elon Musk, Stephen Hawking, and Bill Gates leading the way, I find it strange to see this sentiment echoing across the Internet: predictions of an upcoming AI-induced harmageddon, mockingly referred to as “preemptive extinction panic” (1). In fact, I share that deriding stance, as do many others (2). Yet, as demonstrated in our last newsletter, IAIM researchers are also concerned – and their own similar-sounding headlines predate Hawking’s (3). So have we changed our mind? No. But there is more to this story. Read on.



There are two versions of the harmageddon scenario. One mirrors what many movies so entertainingly portray: A machine magically acquires sentient status and free will, decides to take over the world, and – depending on who the prophet is – enslaves, kills, or turns us humans into paper clips. Such science-fiction inspired scenarios make sweeping assumptions about the motivational foundations of future AGIs (artificial general intelligences), their operational constraints and structural composition, that couldn’t possibly be known at present. We simply cannot take these too seriously – even if Stephen Hawking says so. The other version has several historical parallels, and is much more plausible: As everyone knows, technology can empower the human mind to do good – as well as evil. In that general sense AI is in the same category as nuclear power. But AI is quite different from nuclear weapons in many ways: its power doesn’t lie primarily in threats of its usage; it is much more controllable, and it can be kept secret while still delivering the goods. Think of excessive spying during the cold war, where hundreds of thousands of regular citizens were hired by governments to spy on their own neighbors – soon this could be done with a few PCs and AI software. With future AGIs a single individual may do more than is possible today with a whole army of spies, information gatherers, campaigners, and activists. Like any other technology, AI can be abused at everyone’s expense, shifting power and control in ways that allow gangs and criminals to run rampant, increasing unequal distribution of wealth, and generally escalating the dangers associated with tensions between governments, groups, and nations.



It is this latter scenario that concerns us at IAIM: We want researchers to actively participate in preventing the abuse of knowledge they produce. As the goal of IAIM is to benefit the general public, strengthen industry, and facilitate collaboration, today we are doing something noteworthy on a global scale:



IIIM is announcing its new Ethics Policy for Peaceful R&D. The policy, printed in its entirety in this newsletter, states in no uncertain terms that IIIM's work is to be for the benefit of all. Of course, it is not unethical to create intelligent machines, what matters is how they are used. The policy has the unanimous support from IIIM's employees and Board of Directors. What does that say? It says we really mean it. Being the first AI R&D lab to create and adopt such a policy, we consider it a major achievement, and are very proud to see it instantiated. Details of execution remain to be fleshed out, but the course has been set. Onwards and upwards – for the benefit of all!


REFS

1. Elon Musk, Stephen Hawking warn of artificial intelligence dangers. Mashable, Adario Strange, January 14 2015.
2. The Maverick Nanny With a Dopamine Drip: Debunking Fallacies in the Theory of AI Motivation. R. Loosemore, IEET blog, July 2014. <http://ieet.org/index.php/IEET/more/loosemore20140724>
3. Could Mean the End of Humanity', DV, February 17, 2014.

Preamble

We are proud to announce our brand new Ethics Policy for Peaceful R&D. The policy takes aim at two major threats to societal prosperity and peace. On the one hand, increases in military spending continue throughout the world, including automated weapons development. Justified by “growing terrorist threats”, these actions are themselves resulting in increased use of undue and unjustified force, military and otherwise – the very thing they are aiming to suppress. On the other, the increased possibility – and in many cases clearly documented efforts – of governments wielding advanced technologies to spy on their law-abiding citizens, in numerous ways, and sidestep long-accepted public policy intended to protect private lives from public exposure has gradually become too acceptable. In the coming years and decades artificial intelligence (AI) technologies – and powerful automation in general – has the potential to make these matters significantly worse.

In the US a large part of research funding for AI has come from the military. Since WWII, Japan took a clear-cut stance against military-oriented research in its universities, standing for over half a century as a shining example of how a whole nation could take the pacifist high road. Instead of more countries following its lead, the exact reverse is now happening: Japan is



relaxing these constraints (1), as funding for military activities continues to grow in the U.S., China, and elsewhere. And were it not for the extremely brave actions of a single individual, Edward Snowden, we might still be in the dark about the NSA's pervasive breach of the U.S. constitution, trampling on civil rights that took centuries to establish.

In the past few years the ubiquity of AI systems, such as Apple's Siri, Google's powerful search engine, and IBM's question-answering system Watson, has resulted in a waxing interest in AI across the globe, increasing funding available for such technologies in all its forms. We should expect a speedup, not a status quo or slowdown, of global advances and adaptation of AI technologies, across all industries. It is becoming increasingly important for researchers and laboratories to take a stance on who is to benefit from their R&D efforts – just a few individuals, groups, and governments, or the general people of planet Earth? This is what we are doing today. This is why our Ethics Policy for Peaceful R&D exists. As far as we know, no other R&D laboratory has initiated such a policy.

REF

1. Japan Looks to End Taboo on Military Research at Universities – Government wants to tap best scientists to bolster defenses. By Eric Pfanner and Chieko Tsuneoka, March 24, 2015, 11:02 p.m. ET

IIM'S ETHICS POLICY FOR PEACEFUL R&D:



The Board of Directors of IIM believes that the freedom of researchers to explore and uncover the principles of intelligence, automation, and autonomy, and to recast these as the mechanized runtime principles of man-made computing machinery, is a promising approach for producing advanced software with commercial and public applications, for solving numerous difficult challenges facing humanity, and for answering important questions about the nature of human thought.

A significant part of all past artificial intelligence (AI) research in the world is and has been funded by military authorities, or by funds assigned various militaristic purposes, indicating its importance and application to military operations. A large portion of the world's most advanced AI research is still supported by such funding, as opposed to projects directly and exclusively targeting peaceful civilian purposes. As a result, a large and disconcerting imbalance exists between AI research with a focus on hostile applications and AI research with an explicitly peaceful agenda. Increased funding for military research has a built-in potential to fuel a continual arms race; reducing this imbalance may lessen chances of conflict due to international tension, distrust, unfriendly espionage, terrorism, undue use of military force, and unjust use of power.

Just as AI has the potential to enhance military operations, the utility of AI technology for enabling perpetration of unlawful or generally undemocratic acts is unquestioned. While less obvious at present than the military use of AI and other advanced technologies, the falling cost of computers is likely to make highly advanced automation technology increasingly accessible to anyone who wants it. The potential for all technology of this kind to do harm is therefore increasing.



For these reasons, and as a result of IIIM's sincere goal 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, IIIM's Board of Directors hereby states the Institute's stance on such matters clearly and concisely, by establishing the following Ethical Policy for all current and future activities of IIIM:

1 - IIIM's aim is to advance scientific understanding of the world, and to enable the application of this knowledge for the benefit and betterment of humankind.

2 - IIIM will not undertake any project or activity intended to (2a) cause bodily injury or severe emotional distress to any person, (2b) invade the personal privacy or violate the human rights of any person, as defined by the United Nations Declaration of Human Rights, (2c) be applied to unlawful activities, or (2d) commit or prepare for any act of violence or war.

2.1 - IIIM will not participate in projects for which there exists any reasonable evidence of activities 2a, 2b, 2c, or 2d listed above, whether alone or in collaboration with governments, institutions, companies, organizations, individuals, or groups.

2.2 - IIIM will not accept military funding for its activities.

'Military funding' is defined as any and all funds designated to support the activities of governments, institutions, companies, organizations, and groups, explicitly intended for furthering a military agenda, or to prepare for or commit to any act of war.

2.3 - IIIM will not collaborate with any institution, company, group, or organization whose existence or operation is explicitly, whether in part or in whole, sponsored by military funding as described in 2.2 or controlled by military authorities. For civilian institutions with a history of undertaking military-funded projects a 5-15 rule will be applied: if for the past 5 years 15% or more of their projects were sponsored by such funds, they will not be considered as IIIM collaborators.



IIIM AND CADIA'S AI FESTIVAL

On October 31st the annual AI Festival was held at Reykjavik University. The festival allowed the public to get a glimpse into the future of this fast-growing research field. The Icelandic Institute of Intelligent Machines and RU's CADIA (Center for Analysis & Design of Intelligent Agents) hosted and organized the festival, which was part of Reykjavik University's anniversary celebrations. The festival generated much interest, with over 100 people attending and participating. This year's theme was artificial general intelligence (AGI) and the festival was divided into two main parts, public talks and an exhibition of the latest developments in AI and high-tech in Iceland. Participants got two unique perspectives on these topics, that of researchers developing new innovations and the companies deploying AI and related high-tech solutions with their products and services.



This year's festival was also the launch pad for IIIM and CADIA's new collaborative outreach program to Icelandic startups in the high-tech industry, the High-Tech Highway (Hraðbrautin). Startups attended a meeting where IIIM's and CADIA's research fields were introduced to possibilities for collaboration. This session proved highly successful, with many interested startups voicing their development challenges that IIIM and CADIA could help with, having the in-house knowledge and technologies needed to evaluate, prototype, and test ideas and propel R&D forward.



RU's Sun building was the venue for the AI and high-tech exhibition, where researchers and companies showcased their work, products and solutions to the interested public. The showcase featured a diverse assortment of large and small companies whose products and research involved AI and high-tech solutions.



IIIM AND CADIA'S HÁTÆKNI- HRADBRAUT (HIGH-TECH HIGHWAY):

Opportunity for startups to access expert R&D knowledge, meet future collaborators, and increase their potential for success.

Opportunity for startups to access expert R&D knowledge, meet future collaborators, and increase their potential for success.

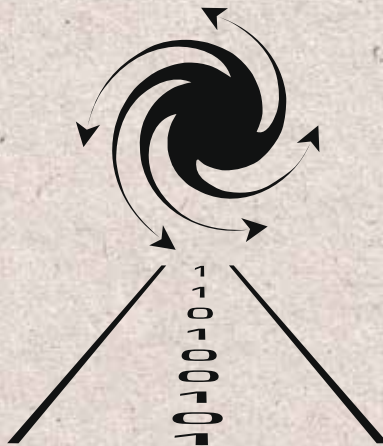
IIIM has joined forces with RU's CADIA, and is offering startups an opportunity to increase their chances of success by contacting the High-Tech Highway (Icelandic: Hátæknihradbrautin).

IIIM and CADIA have worked closely together in the past and this new type of collaboration is particularly suited to help new startups explore cutting-edge possibilities, avoid reinventing the wheel, and get expertise in honing their product plans.

IIIM and CADIA's researchers have extensive knowledge of the following:

- Software Engineering and System Design Development
- Big Data
- Artificial Intelligence
- Simulations of complex systems and processes
- Grants and grant proposal writing
- Founding startups, managing dev teams, talking to investors
- Working with an international network of researchers, research institutes and professionals in AI and high-tech development
- ... and much more.

If you want to know more go to <http://www.iiim.is/hataeknihradbraut/>, <http://www.cadia.ru.is/hataeknihradbraut/> or send us an email: coffee@iiim.is or coffee@cadia.ru.is

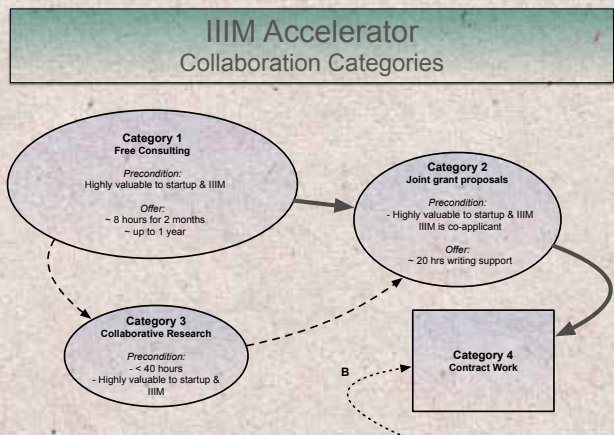


IIIM'S ACCELERATOR / HRAÐALL VITVÉLA- STOFNUNAR

IIIM's Accelerator (IIIM-HTA) is a support system for those who have participated in IIIM and CADIA's High-tech Highway and others who are ready for a close partnership with targeted outcomes. IIIM helps industry by contributing to research and development in many ways, for example:

- Increasing the flow of people, ideas, challenges, chances and research projects to increase efficiency, development of new technology, products and services.
- Assisting startups in understanding the possibilities of technology and consulting on new technologies.
- Our researchers have extensive experiences in both product development and academic research and understand how to improve already developed products, new products and fit the constraints that a real market environment sets.

IIIM-HTA has four main categories of collaboration with industry. Category 1 is the most common for young startups: consulting sessions where the ideas, questions, and challenges identified by the startup are dissected, analyzed, and discussed with members from IIIM and their relevant collaborators. This category is the most common initial category for any IIIM collaborator from industry. A common path for successful collaboration is indicated with the bold arrow, progressing from Cat-1 to Cat-2 to Cat-3. Another common path is indicated by the dotted arrow: Cat-1 to Cat-3 to Cat-4. There is no requirement, however, for any IIIM collaborator from industry to move between categories – a pathway to success is determined on a per-project / per-collaboration basis.



AROUND THE GLOBE

**RECENT
PUBLICATIONS
& TECH REPORTS**

- Bieger, J., Thórisson, K. R. & Garrett, D. (2014). Raising AGI: Tutoring Matters. In B. Goerzel, L. Orseau & J. Snaider (eds.), *Proceedings of Artificial General Intelligence (AGI-14)*, 1-10, Quebec, Canada.
- Garrett, D., Bieger, J. & Thorisson, K. R. (2014). Tunable and Generic Problem Instance Generation for Multi-objective Reinforcement Learning. *Proceedings of the 2014 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning*. Orlando, USA. 2014.
- Garrett, D., Bieger, J., & Thorisson, K. R. (2014). Tunable and Generic Problem Instance Generation for Multi-objective Reinforcement Learning. *Proceedings of the 2014 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning* (to appear). Orlando, USA. 2014.
- Garrett, D. & Thorisson, K. R. (2014). Machine Learning for Improving Adaptivity of Artificial Knees to Environmental Conditions. IIMM Technical Report IIMMTR-2014-09-03.
- Gudjonsson, D. S. (2014). Exploring the Potential Macroeconomic Impacts of Branch Banking Practices. Unpublished Master Thesis, Reykjavik University, Iceland.
- Helgason, H. P., K. R. Thorisson, D. Garrett & E. Nivel (2014). Towards a General Attention Mechanism for Embedded Intelligent Systems. *International Journal of Computer Science and Artificial Intelligence*, 4(1): 1-7.
- Nivel, E., K. R. Thórisson, B. R. Steunebrink, H. Dindo, G. Peluzo, M. Rodriguez, C. Hernandez, D. Ognibene, J. Schmidhuber, R. Sanz, H. P. Helgason & A. Chella (2014). Bounded Seed-AGI. In B. Goerzel, L. Orseau & J. Snaider (eds.), *Proceedings of Artificial General Intelligence (AGI-14)*, 85-96, Quebec, Canada.
- Nivel, E., K. R. Thórisson, B. R. Steunebrink, H. Dindo, G. Peluzo, M. Rodriguez, C. Hernandez, D. Ognibene, J. Schmidhuber, R. Sanz, H. P. Helgason, A. Chella & G. Jonsson (2014). Autonomous Acquisition of Natural Language. In A. P. dos Reis, P. Kommers & P. Isaías (eds.), *Proceedings of the IADIS International Conference on Intelligent Systems & Agents 2014 (ISA-14)*, 58-66, Lisbon, Portugal, July 15-17. Recipient of the Outstanding Paper Award at the Intelligent Systems & Agents conference.
- Valgardsson, G. S., F. Fornari, H. Th. Thórisson & K. R. Thórisson (2014). LivingShadows Developer Documentation v. 1.0. IIMM Technical Report IIMMTR-2014-01-001.
- Garrett, D. (2013). Tunable Instance Generation for Many-Task Reinforcement Learning. Icelandic Institute for Intelligent Machines IIMM Technical Report IIMMTR-2013-01-002.
- Kristjánsson, Th. & K. R. Thórisson (2013). The Ethics of Artificial Intelligence: Risks & Responsibilities. IIMM Technical Report IIMMTR-2013-12-002.
- Mallett, J. (2013). Threadneedle: A simulation framework for exploring the behaviour of modern banking systems. Conference on Developments in Economic Education (DEE), September.

VITVÉLASTOFNUN ÍSLANDS SES

Vitvélastofnun Íslands ses er sjálfseignarstofnun með það höfuðmarkmið að brúa milli iðnaðar og háskólarannsóknar og að hraða nýsköpun í íslenskum hátækniíðnaði. Náð samstarf stofnunarinnar við Tölvunarfræðideild Háskólans í Reykjavík tryggir tengsl við fremstu vísindamenn landsins á helstu tækni sviðum svo sem fræðilegri tölvunarfræði, stærðfræði, verkfræði og gervigreind.

Rannsóknir Vitvélastofnunar stærstum hluta knúna áfram af þörfum iðnaðarinnar og niðurstöðurnar hafa nýtingarmöguleika á mörgum sviðum, s.s. við framleiðslu, tölvuleikjagerð, þjálfun með aðstoð tölvutækni, lífupplýsingafræði, orkukerfum og vélmennastýringu.

Vitvélastofnun leggur áherslu á flýta fyrir og bæta árangur fyrirtæka, breikka sjónvildarhorning þeirra og auka möguleika þeirra að koma hátækni vörum fyrir á markað. Með samstarfsaðilum okkar vinnum við hinitmiðuð verkefni, bættum og aukum samskipti og flæði viðeigandi upplýsinga milli aðila, sækjum um styrki, veitum ráðgjöf og þróum frumgerðir.

IIIM THE ICELANDIC INSTITUTE FOR INTELLIGENT MACHINES

The Icelandic Institute for Intelligent Machines (IIIM) is a non-profit research institute that catalyzes innovation through a focused exchange of ideas, people, projects, and intellectual property. Through close affiliation with Iceland's strongest technological academic department, Reykjavik's School of Computer Science, we bridge the gap between industrial engineering needs and academic research results.

Our work is driven by the needs of industry, and has relevance to a wide range of application areas. To name just a few: computer-based training, bioinformatics, computer games, energy system, virtual and augmented realities, robotics, artificial intelligence, machine learning, and data manipulation, IIIM's software tools, methods, and systems help companies see further into the future, bring high technology to their product lines, and produce more advanced products faster.

CONTACT

IIIM is located on the 2nd floor of Reykjavik University's new Millennium building in Nautholsvik, within unique outdoors areas and near the country's only artificial beach.

Icelandic Institute for Intelligent Machines
Menntavegur 1, Uranus, 2nd fl. IS-101 Reykjavik, Iceland

info@iiim.is

+354.552.1020 (voice)
+354.872.0026 (fax)

