**IOSR Journal of Engineering (IOSR JEN)** **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 18-25

# Super Intelligence Futurescope

Ms. Vikas K. Ravidas[1], Prof. Rashmi Dukhi[2]
*1. Student, Department of Computer Science, GHRIIT, Nagpur, Maharashtra*
*2. Assistant Professor, Department of Computer Science, GHRIIT, Nagpur, Maharashtra*

***Abstract:*** *Studies of super intelligent-level systems have typically posited AI functionality that plays the role of a mind in a rational utility-directed agent, and hence employ an abstraction initially developed as an idealized model of human decision makers. Today, developments in AI technology highlight intelligent systems that are quite unlike minds, and provide a basis for a different approach to understanding them: Today, we can consider how AI systems are produced (through the work of research and development), what they do (broadly, provide services by performing tasks), and what they will enable (including incremental yet potentially thorough automation of human tasks). Because tasks subject to automation include the tasks that comprise AI research and development, current trends in the field promise accelerating AI-enabled advances in AI technology itself, potentially leading to asymptotically recursive improvement of AI technologies in distributed systems, a prospect that contrasts sharply with the vision of self-improvement internal to opaque, unitary agents. The concept of comprehensive AI services (CAIS) provides a model of flexible, general intelligence in which agents are a class of service-providing products, rather than a natural or necessary engine of progress in themselves. Ramifications of the CAIS model reframe not only prospects for an intelligence explosion and the nature of advanced machine intelligence, but also the relationship between goals and intelligence, the problem of harnessing advanced AI to broad, challenging problems, and fundamental considerations in AI safety and strategy. Perhaps surprisingly, strongly self-modifying agents lose their instrumental value even as their implementation becomes more accessible, while the likely context for the emergence of such agents becomes a world already in possession of general superintelligent-level capabilities.*

***Keywords:*** *Artificial Intelligence(AI), Center for Artificial IntelligenceinSociety(CAIS)*

## I. Introduction

Responsible development of AI technologies can provide an increasingly comprehensive range of superintelligent-level AI services including the service of developing new services and can thereby deliver the value of general-purpose AI while avoiding the risks associated with self-modifying AI agents.The emerging trajectory of AI development reframes AI prospects. Ongoing automation of AI R&D tasks, in conjunction with the expansion of AI services, suggests a tractable, non-agent-centric model of recursive AI technology improvement that can implement general intelligence in the form of comprehensive AI services (CAIS), a model that includes the service of developing new services. The CAIS model—which scales to superintelligent-level capabilities—follows software engineering practice in abstracting functionality from implementation while maintaining the familiar distinction between application systems and development processes. Language translation exemplifies a service that could incorporate broad, superintelligent-level world knowledge while avoiding classic AI-safety challenges both in development and in application. Broad world knowledge could likewise support predictive models of human concerns and (dis)approval, providing safe, potentially superintelligent-level mechanisms applicable to problems of AI alignment. Taken as a whole, the R&D-automation/CAIS model reframes prospects for the development and application of superintelligence, placing prospective AGI agents in the context of a broader range of intelligent systems while attenuating their marginal instrumental value

## II. Methodology

### AI development reframes AI prospects:

Past, present, and projected developments in AI technology can inform our understanding of prospects for superintelligent-level capabilities, providing a concrete anchor that complements abstract models of potential AI systems. A development-oriented perspective highlights path-dependent considerations in assessing potential risks, risk-mitigation measures, and safety-oriented 17 research strategies. The current trajectory of AI development points to asymptotically recursive automation of AI R&D that can enable the emergence of general, asymptotically comprehensive AI services (CAIS). In the R&Dautomation/CAIS model, recursive improvement and general AI capabilities need not be embodied in systems that act as AGI agents

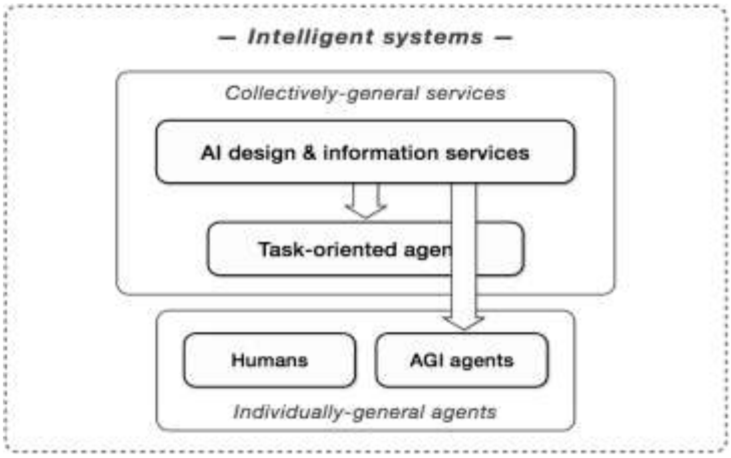
— Intelligent systems —
Collectively-general services
AI design & information services
Task-oriented agen
Humans
AGI agents
Individually-general agents

***Figure 1: Classes of intelligent systems***

**A *technology-centered* model of recursive improvement:**Technology improvement proceeds through research and development, a transparentprocess that exposes component tasks to inspection, refactoring, andincremental automation.1 If we take advanced AI seriously, then accelerating,asymptotically-complete R&D automation is a natural consequence:

- By hypothesis, advances in AI will enable incremental automation andspeedup of all human tasks.
- As-yet unautomated AI R&D tasks are human tasks, hence subject toincremental automation and speedup.
- Therefore, advances in AI will enable incremental automation andspeedup of all AI R&D tasks.

Today we see automation and acceleration of an increasing range of AI R&Dtasks, enabled by the application of both conventional software tools and technologiesin the AI spectrum. Past and recent developments in the automationof deep-learning R&D tasks include:

- Diverse mechanisms embodied in NN toolkits and infrastructures
- Black-box and gradient-free optimization for NN hyperparameter search(Jaderberg et al. 2017)
- RL search and discovery of superior NN gradient-descent algorithms(Bello et al. 2017)
- RL search and discovery of superior NN cells and architectures (Zophet al. 2017)

Today, automation of search and discovery (a field that overlaps with “metalearning”)requires human definition of search spaces, and we can expectthat the *definition of new search spaces*—as well as fundamental innovationsin architectures, optimization methods, and the definition and constructionof tasks—will remain dependent on human insight for some time to come.Nonetheless, increasing automation of even relatively narrow search and discoverycould greatly accelerate the implementation and testing of advances based on human insights, as well as their subsequent integration with othercomponents of the AI technology base. Exploring roles for new components(including algorithms, loss functions, and training methods) can be routine,yet important: as Geoff Hinton has remarked, “A bunch of slightly new ideasthat play well together can have a big impact”.Focusing exclusively on relatively distant prospects for *full* automationwould distract attention from the potential impact of incremental researchautomation in accelerating automation itself.

**A*service-centered* model of generalintelligence:**AI deployment today is dominated by AI services such as language translation,image recognition, speech recognition, internet search, and a host of servicesburied within other services. Indeed, corporations that provide cloud computingnow actively promote the concept of “AI as a service” to other corporations.Even applications of AI within autonomous systems (*e.g.,* self-driving vehicles) can be regarded as providing services (planning, perception, guidance) toother system components.R&D automationcan itself be conceptualized as a set of services thatdirectly or indirectly enable the implementation of new AI services. Viewingservice development through the lens of R&D automation, tasks for advanced AI include:

- Modeling human concerns
- Interpreting human requests
- Suggesting implementations
- Requesting clarifications
- Developing and testing systems
- Monitoring deployed systems
- Assessing feedback from users
- Upgrading and testing systems

CAIS functionality, which includes the service of developing stable, taskorientedAI agents, subsumes the instrumental functionality of proposedself-transforming AGI agents, and can present that functionality in a form that better fits the established conceptual frameworks of business innovationand software engineering.

**The services model abstracts *functionality* from*implementation:***Describing AI systems in terms of functionalbehaviors (“services”) aligns withconcepts that have proved critical in software systems development. Theseinclude separation of concerns, functional abstraction, data abstraction, encapsulation,and modularity, including the use of client/server architectures—aset of mechanisms and design patterns that support effective program design,analysis, composition, reuse, and overall robustness.Abstraction of functionality from implementation can be seen as a figuregroundreversal in systems analysis. Rather than considering a complexsystem and asking how it will behave, one considers a behavior and asks howit can be implemented. Desired behaviors can be described as services, andexperience shows that complex services can be provided by combinationsof more specialized service providers, some of which provide the service of
aggregating and coordinating other service providers.

**Model distinguishes *development* from*functionality:***The AI-services model maintains the distinction betweenAI *development* andAI *functionality*. In the development-automation model of advanced AI services,
stable systems build stable systems, avoiding both the difficulties and potentialdangers of building systems subject to open-ended self-transformationand potential instability.Separating development from application has evident advantages. For one,task-focused applications need not themselves incorporate an AI-developmentapparatusthere is little reason to think that a system that provides onlinelanguage translation or aerospace engineering design services should in additionbe burdened with the tasks of an AI developer. Conversely, large resourcesof information, computation, and time can be dedicated to AI development,far beyond those required to perform a typical service. Likewise, in ongoingservice application and upgrade, aggregating information from multiple deployedsystems can provide decisive advantages to centralized development(for example, by enabling development systems for self-driving cars to learnfrom millions of miles of car-experience per day). Perhaps most important,stable products developed for specific purposes by a dedicated developmentprocess lend themselves to extensive pre-deployment testing and validation.

**Language translation exemplifies a safe, potentiallysuperintelligent service:** Language translation provides an example of a service best provided bysuperintelligent-level systems with broad world knowledge. Translationof written language maps input text to output text, a bounded, episodic,sequence-to-sequence task. Training on indefinitely large and broad textcorpora could improve translation quality, as could deep knowledge of psychology,philosophy, history, geophysics, chemistry, and engineering. Effectiveoptimization of a translation system for an objective that weights both qualityand efficiency would focus computation solely on the application of thisknowledge to translation.The process of *developing* language translation systems is itself a servicethat can be formulated as an episodic task, and as with translation itself,effective optimization of translation-development systems for both qualityand efficiency would focuscomputation solely on that task.There is little to be gained by modeling stable, episodic service-providersas rational agents that optimize a utility function over future states of theworld, hence a range of concerns involving utility maximization (to say nothingof self-transformation) can be avoided across a range of tasks. Evensuperintelligent-level world knowledge and modeling capacity need not initself lead to strategic behavior.

**Human (dis)approval can aid AI goalalignment:** As noted by Stuart Russell, written (and other) corpora provide a rich sourceof information about human opinions regarding actions and their effects;intelligent systems could apply this information in developing predictivemodels of human approval, disapproval, and disagreement. Potential trainingresources for models of human approval include existing corpora of text
and video, which reflect millions of person-years of both real and imaginedactions, events, and human responses; these corpora include news, history,fiction, science fiction, advice columns, law, philosophy, and more, and couldbe augmented and updated with the results of crowd-sourced challengesstructured to probe model boundaries.Predictive models of human evaluations could provide strong priors andcommon-sense constraints to guide both the implementation and actions of AIservices, including strategic advisory services to powerful actors. Predictivemodels are not themselves rational agents, yet models of this kind couldcontribute to the solution of agent-centered safety concerns. In this connection,separation of development from application can insulate such models fromperverse feedback loops involving self-modification.

**CAIS model reframes prospects forsuperintelligence:**
From a broad perspective, the R&D-automation/CAIS model:

- Distinguishes recursive technology improvement from self-improvingagents

- Shows how incremental automation of AI R&D can yield recursiveimprovement
- Presents a model of general intelligence centered on services rather thansystems
- Suggests that AGI agents are not necessary to achieve instrumental goals
- Suggests that high-level AI services would precede potential AGI agents
- Suggests potential applications of high-level AI services to general AIsafety

For the near term, the R&D-automation/CAIS model:

- Highlights opportunities for safety-oriented differential technology development
- Highlights AI R&D automation as a leading indicator of technologyacceleration
- Suggests rebalancing AI research portfolios toward AI-enabled R&Dautomation

Today, we see strong trends toward greater AI R&D automation and broaderAI services. We can expect these trends to continue, potentially bridging thegap between current and superintelligent-level AI capabilities. Realistic, pathdependentscenarios for the emergence of superintelligent-level AI capabilitiesshould treat these trends both as an anchor for projections and as a prospectivecontext for trend-breaking developments.

## III. Performance Evaluation

➢ **This document outlines topics, questions, and propositions that address:**

1. Prospects for an intelligence explosion
2. The nature of advanced machine intelligence
3. The relationship between goals and intelligence
4. The problem of using and controlling advanced AI
5. Near- and long-term considerations in AI safety and strategy

The questions and propositions below reference sections of this document thatexplore key topics in more depth. From the perspective of AI safety concerns,this document offers support for several currently-controversial propositionsregarding artificial general intelligence:

- That AGI agents have no natural role in developing general AI capabilities.
- That AGI agents would offer no unique and substantial value in providinggeneral AI services.
- That AI-based security services could safely constrain subsequent AGIagents, even if these operate at a superintelligent level.

➢ **Reframing prospects for an intelligence explosion**

- Does recursive improvement imply self-transforming agents?

Ongoing automation of tasks in AI R&D suggests a model of asymptotically recursive technology improvement that scales to superintelligent-level (SIlevel) systems. In the R&D-automation model, recursive improvement is systemic, not internal to distinct systems or agents. The model is fully generic: It requires neither assumptions regarding the content of AI technologies, nor assumptions regarding the pace or sequence of automation of specific R&D tasks. Classic self-transforming AGI agents would be strictly more difficult to implement, hence are not on the short path to an intelligence explosion.

- Can fast recursive improvement be controlled and managed?

Recursive improvement of basic AI technologies would apply allocated machine resources to the development of increasingly functional building blocks for AI applications (better algorithms, architectures, training methods, etc.); basic technology development of this sort could be open-ended, recursive, and fast, yet non-problematic. Deployed AI applications call for careful management, but applications stand outside the inner loop of basic-technology improvement.

- Would general learning algorithms produce systems with general competence?

The application of an idealized, fully general learning algorithm would enable but not entail the learning of any particular competence. Time, information, and resource constraints are incompatible with universal competence, regardless of ab initio learning capacity

➢ **Reframing the nature of advanced machine intelligence**

- Is human learning an appropriate model for AI development?

Action, experience, and learning are typically decoupled in AI development: Action and experience are aggregated, not tied to distinct individuals, and the machine analogues of cognitive change can be profound during system development, yet absent in applications. As we see in AI technology today, learning algorithms can be applied to produce and upgrade systems that do not themselves embody those algorithms. Accordingly, using human learning and action as a model for AI development and application can be profoundly misleading.

- Does stronger optimization imply greater capability?

Because optimization for a task focuses capabilities on that task, strong optimization of a system acts as a strong constraint; in general, optimization does not extend the scope of a task or increase the resources employed to perform it. System optimization typically tends to reduce resource consumption, increase throughput, and improve the quality of results.

- Do broad knowledge and deep world models imply broad AI capabilities?

Language translation systems show that safe, stable, high-quality task performance can be compatible with (and even require) broad and deep knowledge about the world. The underlying principle generalizes to a wide range of tasks.

- Must we model SI-level systems as rational, utility-maximizing agents?

The concept of rational, utility-maximizing agents was developed as an idealized model of human decision makers, and hence is inherently (though abstractly) anthropomorphic. Utility-maximizing agents may be intelligent systems, but intelligent systems (and in particular, systems of agents) need not be utility-maximizing agents.

- Must we model SI-level systems as unitary and opaque?

Externally-determined features of AI components (including their development histories, computational resources, communication channels, and degree of mutability) can enable structured design and functional transparency, even if the components themselves employ opaque algorithms and representations.

➢ Reframing the relationship between goals and intelligence

- What does the orthogonality thesis imply for the generality of convergent instrumental goals?

Intelligent systems optimized to perform bounded tasks (in particular, episodic tasks with a bounded time horizon) need not be agents with open-ended goals that call for self preservation, cognitive enhancement, resource acquisition, and so on; by Bostrom's orthogonality thesis, this holds true regardless of the level of intelligence applied to those tasks.

- How broad is the basin of attraction for convergent instrumental goals?

Instrumental goals are closely linked to final goals of indefinite scope that concern the indefinite future. Societies, organizations, and (in some applications) high-level AI agents may be drawn toward convergent instrumental goals, but high-level intelligence per se does not place AI systems within this basin of attraction, even if applied to broad problems that are themselves long-term.

➢ Reframing the problem of using and controlling advanced AI

- Would the ability to implement potentially-risky self-transforming agents strongly motivate their development?

If future AI technologies could implement potentially-risky, self-transforming AGI agents, then similar, more accessible technologies could more easily be applied to implement open, comprehensive AI services. Because the service of providing new services subsumes the proposed instrumental value of self-transforming agents, the incentives to implement potentially-risky self-transforming agents appear to be remarkably small.

- Can we architect safe, superintelligent-level design and planning services?

Consideration of concrete task structures and corresponding services suggests that SI-level AI systems can safely converse with humans, perform creative search, and propose designs for systems to be implemented and deployed in the world. Systems that provide design and planning services can be optimized to provide advice without optimizing to manipulate human acceptance of that advice.

## IV. Major Categories

Artificial intelligence is interesting for millions of people around the world. We want to have robots clean our house, talk to us, and support. Still, we often think about types of artificial intelligence as about mighty robots that will take over most of our jobs. Can this happen in the near future? Not likely. However, there are already many awesome kinds of artificial intelligence.

Let's start with major artificial intelligence categories:

- ANI
- AGI
- ASI

ANI stands for Artificial Narrow Intelligence. It's often called Weak AI. This AI is considered to be weak because it specializes in only one category. AGI, Artificial General Intelligence is also called Strong AI or Human-Level AI. As you may understand from the name, AGI has the same capabilities as a human. ASI, Artificial Super intelligence, is the type of intelligence that is smarter than humans.

Nowadays the world is running on ANI. AGI may be created in the near future, while ASI – in the distant future. However, you may check out ANI examples, because they are everywhere.

**ANI examples:**

Have you got a car? Look for self-driving cars. For example, check out Google autonomous electric cars that are controlled by Google Chauffeur software. There are already functional prototypes. These driverless cars don't have steering wheels or pedals. Being equipped with LIDAR system and a Velodyne 64-beam laser, Google cars generate 3D maps of the environment and use them for driving.

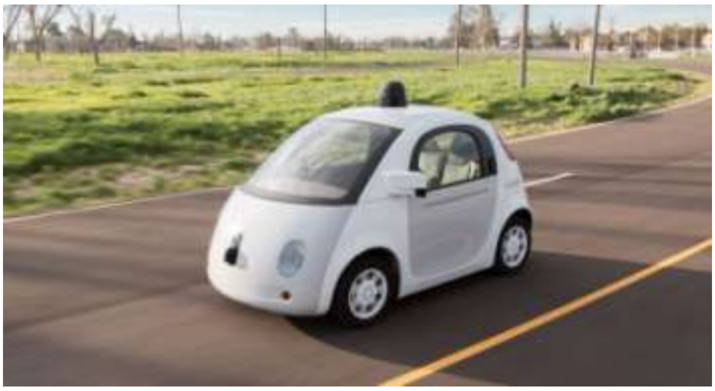

***Figure2. Self Driving Car***

According to the results presented in June 2015, Google self-driving vehicles have driven more than 1,000,000 mi. Besides, they've encountered 180 million other vehicles and 200,000 stop signs. The testing speed isn't higher than 25 mph. Besides, there is always a safety driver aboard. However, on February 2016 a self-driving car struck a bus while trying to avoid sandbags. There were incidents of other types. However, most of them happened because of other drivers making mistakes.Google still has to work on some issues. For example, this type of cars has problems with identifying trash and light debris on the road, spotting police officers or other people who signal the car to stop. However, Google plans to fix these issues by 2020.

.

Mail filters are another type of artificial intelligence. Using artificial neural network, they spot and flag special messages. They identify the source of letters and keep the spam box level low. Of course, these filters aren't perfect. How many times have you missed an important email because it got into a spam box? It happened to me more than 3 times. That's why I'm happy that companies are working on improving this service. Example, Google is changing its detection algorithms. The next type of artificial narrow intelligence is speech recognition software.

**Are robots ANI or AGI?**

Robots are also types of artificial intelligence. As most of them have only several functions, they are kinds of ANI. That means they are good at something but they can't be perfect assistants that are cleverer than humans.In order to check whether some types of artificial intelligence fall into ANI or AGI, scientists use several tests.

- The first test is called the Turing test. It tests the ability of a computer or a machine to exhibit behavior similar to that of a human. For example, evaluators had to test conversations between a human and a machine. If a machine has convinced the evaluator 70% of the conversation time, it has passed the test.The Turing test has many interpretations. One of them is used for choosing the winner of the Loebner Prize. If a program or application can fool half of the judges into considering it to be a human, than it will win the first prize. The conversation should last for 30 minutes.
- The second variant is easier to understand. It's called a coffee test. A robot's goal is to go into an average home, identify coffee, and figure out how to make it properly. If a robot can do that effectively, it may be perceived as a type of AGI.

- The third test you will like is the robot college student test. If a robot can enroll into the university and take classes the same way people do, it is a type of artificial general intelligence. It'd be interesting to have a robot as your course mate.

***Figure3 : Robot are ANI or AGI***

The next test is connected with employment. A robot will be perceived as a type of AGI if it can pass vocational tests. Additional tasks include driving and writing exams. All these tests seem to be interesting and effective. These robots must be clever learners that can make decisions, answer questions, make conclusions, and use different approaches for solving various types of problems. The robots we have now aren't AGI. However, they are effective in carrying out their tasks.

For example, there is a personal home health robot Pillo that helps people manage their health. Pillo looks modern and stylish. It can answer wellness and health questions. Pillo may contact healthcare professionals. Besides, it stores and dispenses vitamins and medications. If needed, this health robot may order refills. With voice and facial recognition technology, Pillo can recognize all family members and remind everyone about the pills or vitamins they need. This robot is a nice companion that cares about the health of all family members. Now it's only a prototype. I'm eager to watch this robot functioning in everyday life.The next robot I'd like to check out is Alpha 2, the first humanoid robot. In future, this robot must carry out a bunch of functions including tutoring, interpreting, smart home management. It's a type of artificial intelligence that will function as a weatherman and storyteller, a home office assistant and an entertainer. It will be able to dance with you, teach you some yoga poses, and web search for all questions you have. Alpha 2 is already in stores. Still, it doesn't have all the previously mentioned features now. Developers are working on it to make effective robots a reality as soon as possible. As you see, there are three types of artificial intelligence: ANI, AGI, and ASI. ANI is already here. Scientists, developers, and other people work on it to take ANI to the next level and bring effective AGI into the world.

**What does *Artificial Superintelligence (ASI)*mean?**

Artificial superintelligence is a term referring to the time when the capability of computers will surpass humans. "Artificial intelligence," which has been much used since the 1970s, refers to the ability of computers to mimic human thought. Artificial superintelligence goes a step beyond, and posits a world in which a computer's cognitive ability is superior to a human's.

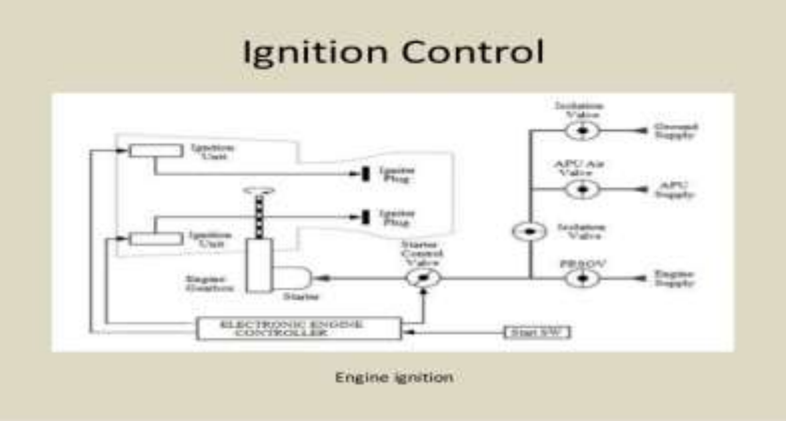

**Figure 4:** ***ASI  (engine ignition control )***

**Techopedia explains *Artificial Superintelligence (ASI)***

Most experts would agree that societies have not yet reached the point of artificial superintelligence. In fact, engineers and scientists are still trying to reach a point that would be considered full artificial intelligence, where a computer could be said to have the same cognitive capacity as a human. Although there have been developments like IBM's Watson supercomputer beating human players at Jeopardy, and assistive devices like Siri engaging in primitive conversation with people, there is still no computer that can really simulate the breadth of knowledge and cognitive ability that a fully developed adult human has. The Turing test, developed decades ago, is still used to talk about whether computers can come close to simulating human conversation and thought, or whether they can trick other people into thinking that a communicating computer is actually a human. However, there is a lot of theory that anticipates artificial superintelligence coming sooner rather than later. Using examples like Moore's law, which predicts an ever-increasing density of transistors, experts talk about singularity and the exponential growth of technology, in which full artificial intelligence could manifest within a number of years,and artificial superintelligence could exist in the 21st century

## V. Conclusion

We can expect to see AI-enabled automation of AI research and development continue to accelerate, both leveraging and narrowing the scope of human insights required for progress in AI technologies. Asymptotically-recursive improvement of AI technologies can scale to a superintelligent level, supporting the development of a fully-general range of high-level AI services that includes the service of developing new services in response to human demand. Because general AI-development capabilities do not imply general capabilities in any particular system or agent, classic AGI agents would be potential products of SI-level AI development capabilities, not a path to uniquely valuable functionality. The development of diverse, high-level AI services also offers opportunities for safety-relevant differential technology development, including the development of common-sense predictive models of human concerns that can be applied to improve the value and safety of AI services and AI agents. The R&D-automation/AI-services model suggests that conventional AI risks (e.g., failures, abuse, and economic disruption) are apt to arrive more swiftly than expected, and perhaps in more acute forms. While this model suggests that extreme AI risks may be relatively avoidable, it also emphasizes that such risks could arise more quickly than expected. In this context, agent-oriented studies of AI safety can both expand the scope of safe agent applications and improve our understanding of the conditions for risk.

## VI. References

[1]. Halal, William E. "TechCast Article Series: The Automation of Thought" (PDF). Archived from the original (PDF) on 6 June 2013.

[2]. Aleksander, Igor (1996), Impossible Minds, World Scientific Publishing Company, ISBN 978-1-86094-036-1

[3]. Omohundro, Steve (2008), The Nature of Self-Improving Artificial Intelligence, presented and distributed at the 2007 Singularity Summit, San Francisco, CA.

[4]. Johnson, Mark (1987), The body in the mind, Chicago, ISBN 978-0-226-40317-5

[5]. Kurzweil, Ray (2005), The Singularity is Near, Viking Press

[6]. Lighthill, Professor Sir James (1973), "Artificial Intelligence: A General Survey", Artificial Intelligence: a paper symposium, Science Research Council

[7]. Luger, George; Stubblefield, William (2004), Artificial Intelligence: Structures and Strategies for Complex Problem Solving (5th ed.), The Benjamin/Cummings Publishing Company, Inc., p. 720, ISBN 978-0-8053-4780-7

[8]. McCarthy, John (October 2007), "From here to human-level AI", Artificial Intelligence, **171** (18): 1174–1182, doi:10.1016/j.artint.2007.10.009.

[9]. McCorduck, Pamela (2004), Machines Who Think (2nd ed.), Natick, MA: A. K. Peters, Ltd., ISBN 1-56881-205-1

[10]. Moravec, Hans (1976), The Role of Raw Power in Intelligence