

**Samu Kumpulainen**

# **Artificial General Intelligence - a systematic mapping study**

Master's Thesis in Information Technology

January 20, 2021

University of Jyväskylä

Faculty of Information Technology

**Author:** Samu Kumpulainen

**Contact information:** samu.p.kumpulainen@student.jyu.fi

**Supervisor:** Vagan Terziyan

**Title:** Artificial General Intelligence - a systematic mapping study

**Työn nimi:** Yleistekoäly - systemaattinen kirjallisuuskartoitus

**Project:** Master's Thesis

**Study line:** Mathematical Information Technology

**Page count:** 37+0

**Abstract:** In this thesis, a systematic mapping study is performed on the field of artificial general intelligence. The goal of the study is to gain insight about the recent developments in the study field. This includes the focus points of the current research, possible research gaps, and how the research itself is conducted. TODO: more accurate, proper abstract

**Keywords:** Master's Theses, AGI, AI, artificial intelligence, systematic literature mapping, mapping study

**Suomenkielinen tiivistelmä:** TODO: translate when abstract done

**Avainsanat:** Pro Gradu, tutkielma, AI, tekoäly

## List of Figures

Figure 1. Process model (Petersen et al. 2008).....	9
Figure 2. Article distribution between the publication forums .....	20
Figure 3. Yearly articles by publication forum .....	21
Figure 4. Topic category frequencies .....	24
Figure 5. Topic's relations to each other .....	25
Figure 6. Bubble plot showing article distribution between topics and article types .....	26

## List of Tables

Table 1. Search phrases .....	15
Table 2. Publication forums and their results .....	17
Table 3. Article classification by Wieringa et al. (2006) .....	18
Table 4. Emergent topic categories .....	22

# Contents

1	INTRODUCTION .....	1
2	ARTIFICIAL GENERAL INTELLIGENCE.....	2
2.1	History of Artificial Intelligence.....	2
2.2	Definition .....	5
3	SYSTEMATIC LITERATURE MAPPING PROCESS.....	8
3.1	Research method description .....	8
3.2	Differences with systematic literature reviews.....	11
3.3	Background/reasoning behind method choice .....	12
4	CONDUCTING THE LITERATURE MAPPING .....	13
4.1	Research questions .....	13
4.2	The material search.....	14
4.2.1	Search phrases .....	15
4.3	Choosing papers for inclusion .....	15
4.4	Keywording.....	17
4.5	Data extraction and mapping .....	17
4.6	Source material control .....	18
5	RESULTS AND ANALYSIS .....	19
5.1	Publications years and venues .....	19
5.2	Common research topics .....	21
5.3	Types of AGI research .....	25
5.4	TODO: Where to put this: Validity threats? .....	27
5.5	TODO: Where to put this: Possible continuation research.....	27
6	CONCLUSION .....	28
	BIBLIOGRAPHY .....	29

# 1 Introduction

The thesis will be a systematic research mapping on the field of Artificial General Intelligence (AGI). The goal of the thesis is to identify the themes and subfields of AGI research in recent years, what is being researched recently, and what kind of gaps exist on the field. For a while the AGI field was not so active and the more specific approaches, 'narrow AI', grew in popularity. Recently, however, the wider, more general artificial intelligence has been regaining interest. This kind of mapping study would be needed as the research field is complex and there is no clear presentation of the current trends and focal points. Creating this kind of overview would be a valuable asset for future research, as it would enable focusing the research on areas less ventured. Furthermore, if an interesting subtopic comes up during the process of mapping, more focus may be directed towards that in form of more traditional systematic literature review. This option is left for further consideration.

What each chapter is about, how the thesis is structured etc.

## 2 Artificial General Intelligence

TODO: Here some chapter introducing text

### 2.1 History of Artificial Intelligence

Even though the idea of autonomous machinery has been around since the ancient Greek (), AI's origins are set around in the 1940s. At the time, American science fiction author Isaac Asimov wrote numerous novels and short stories about conscious robots and technology's relation to humankind. His work has inspired countless people in the field's of AI and computer science since then (Haenlein and Kaplan July 2019). Also in the 1940s, mathematician Alan Turing's work on Britain's code breaking efforts lead to the creation of first electromechanical computer, The Bombe (Haenlein and Kaplan July 2019). Turing later gave lectures and wrote an article titled "*Computing Machinery and Intelligence*" (1950), in which he presented several ideas later prevalent in AI field, including the "Imitation game", a test to measure the intelligence of a machine (Russell and Norvig 2009). This later became well known as the Turing test.

The term Artificial Intelligence was coined in 1956 during a two-month workshop *Dartmouth Summer Research Project on Artificial Intelligence*, organized by John McCarthy and Marvin Minsky (Haenlein and Kaplan July 2019). The participants of the workshop would later become the most prominent figures of AI research. During DSRPAI two researchers, Allen Newell and Herbert Simon presented Logic Theorist, their existing reasoning program, capable of proving multiple mathematical theorems (Russell and Norvig 2009). Based on this work the two later created General Problem Solver, GPS, which could solve simple puzzles like Towers of Hanoi using human like recursive approach (Newell, Shaw, and Simon 1959). The early days of AI research produced many similar results in different areas. IBM's Arthur Samuel created AI programs that learned to play checkers at a strong amateur level (Russell and Norvig 2009). John McCarthy's 1958 paper titled "Programs with common sense", describes Advice Taker, a complete but hypothetical AI system with general knowledge about the world and deductive processes to manipulate it. The paper is still thought to

be relevant today. McCarthy's system was able to acquire new skills in previously unknown areas without being reprogrammed.

During these years also work on the neural networks started to gain interest. The initial work of McCulloch and Pitts (1943), later demonstrated by Hebb (Russell and Norvig 2009), showed that a neural network is capable of learning. In 1960s Rosenblatt's work on perceptrons and Widrow and Hoff's LMS algorithm were some of the biggest advances in the area (Widrow and Lehr 1995). The next great discovery that would propel the neural networks into the focal point of AI research would happen in the mid-1980s when the backpropagation algorithm originally presented by Bryson and Ho in 1969 was rediscovered by multiple independent groups (Russell and Norvig 2009). Backpropagation is one of the most widely used algorithms for training neural networks these days for its relative power and simplicity (Rumelhart et al. 1995).

History of artificial intelligence contains occasional periods of reduced interest and funding. These so called "AI winters" are a result of high expectations collapsing under criticism. First period that can be considered an AI winter started in the 1970s, and Russell and Norvig (2009) present the following possible reasons for it: Firstly, the early programs knew nothing about their context, and solved the problems via syntactic manipulations. This was especially apparent on machine translation projects. As a language cannot be fully understood without knowing the full context of the sentences and other nuances of the language, accurate translation proved to be a difficult task. Failed translation efforts lead to funding cuts in the US. Second difficulty pointed out by (Russell and Norvig 2009) was the sheer complexity of the target problems. As the early AI programs were focused on simple tasks, finding a solution by trial and error was possible in practice. But as the problems became more complex, "combinatorial explosion" issue became more apparent. The issue was also discussed in British scientist James Lighthill's report on the state of the AI (1973). The report is considered to be one of the main reasons why the British government decided to cut all AI funding in all but two universities. Lastly, the limitations of the data structures used in AI field, such as perceptrons, restricted the capabilities of the solutions. According to Russell and Norvig (2009) this lead to funding cuts also in the neural network research.

During and after the first AI winter, there was a considerable amount of research relating

to expert systems (Russell and Norvig 2009). These systems perform their tasks in a way similar to human experts in the specific, narrow domain, relying on a knowledge encoded into a set of rules (Myers 1986). This style of AI research was inspired by the success of DENDRAL (Buchanan, Feigenbaum, and Lederberg 1968), a system developed at Stanford by Ed Feigenbaum, Bruce Buchanan and Joshua Lederberg. DENDRAL's purpose was to use data from mass spectrometer to infer the structure of a given molecule. MYCIN, developed in the 1970s (Shortliffe et al. 1975), incorporated domain knowledge acquired through expert interviews, with the uncertainty of medical evaluation taken into account via certainty factors (Russell and Norvig 2009).

Expert systems gained commercial interest, leading to increased research and adoption in the industry. Government investments in Japan lead to increased funding in United States and Britain, leading to an AI boom in the 1980s (Russell and Norvig 2009). After the boom, at the end of the 1980s, the second AI winter arrived. Participation in AI conferences dropped, several of the new AI companies met their end, as did the AI research divisions in larger hardware and software companies (Nilsson 2009). The imminent burst of the bubble was foreseen by several leading researchers, but their warnings didn't have considerable effect (Nilsson 2009).

According to Russell and Norvig (2009), around this time the AI field started to adopt the scientific method. This means the earlier ways of proposing completely new theories based on vague evidence or oversimplified examples have been replaced by basis on existing theories, repeatable experiments, and real-world examples. This newly discovered open-mindedness then lead to a complete new ways of looking at the AI research. AI solutions based on existing theories, such as speech recognition based on hidden Markov models, enables the researchers to build on the rigorous mathematical theory behind it (Russell and Norvig 2009). Work of Judea Pearl (1988) and Peter Cheeceman (1985) on the probabilistic reasoning lead to it being accepted back into the AI field. Later Pearl's Bayesian networks have been used to handle uncertainty in AI problems. They are graphical models that join probabilistic information and dependencies to events, enabling inference using probabilistic methods (Goertzel and Pennachin 2007).

In the 21st century artificial intelligence research has been steadily growing. According to



(Liu et al. 2018), not only has the amount of publications in the field been increasing, but also the collaboration between researchers. The study also deduces that that AI has become more open-minded and popular, as the rate of self-references is reducing. One reason for the rising popularity on the field is the success that narrow AI solutions have presented in multitude of problems. For example, in classical game of Go, program called AlphaGo developed by Google-owned DeepMind, defeated the world champion Lee Sedol in 2015 (Silver et al. 2016). Due to Go's computationally complex nature this was a impressive feat previously thought impossible. Later DeepMind developed even more advanced versions of AlphaGo, called AlphaGo Zero, and generalized AlphaZero, which could even play Shogi and Chess on superhuman level (Silver et al. 2018).

Recent years majority of the field has been focusing on the narrow AI approaches (Goertzel and Pennachin 2007). However, the interest in the classical, strong AI has also been increasing. This can be seen in the publications from many influential AI researchers. Authors like John McCarthy (2007), Nils Nilsson (2005) and Marvin Minsky (2007) have voiced their opinions that efforts to create a more general AI should be pursued. There are several terms used regarding these efforts. **Human-level Artificial Intelligence** (HLAI) aims to reach "human-level intelligence" and common sense, a goal that according to Marvin Minsky (2004) can be reached by not using any single method, but a combination of different resources and methods. Similar term is **Artificial General Intelligence** (AGI) presented by Ben Goertzel and Casio Pennachin (2007). The goal of AGI is similar to HLA, to create an AI system that can express general intelligence instead of being locked into a single domain. On the next chapter this general approach is presented in more detail, as it is the focus of this thesis.

## 2.2 Definition

In order to be able to define AGI, or artificial intelligence in general, one must first consider the definitions of intelligence in general. There exists many different definitions, in many different branches of science. Legg and Hutter (2007) list over 60 definitions collected from various academic sources. These include, for example, *"the general mental ability involved in calculating, reasoning, perceiving relationships and analogies, learning quickly, storing*

*and retrieving information, using language fluently, classifying, generalizing, and adjusting to new situations.*" (Columbia Encyclopedia, sixth edition, 2006), *"that facet of mind underlying our capacity to think, to solve novel problems, to reason and to have knowledge of the world"* (Anderson 2006), and *"Intelligence is the ability for an information processing system to adapt to its environment with insufficient knowledge and resources."* (Wang 1995). Based on the aforementioned collection of definitions, Legg and Hutter (2007) have formed the following definition: *"Intelligence measures an agent's ability to achieve goals in a wide range of environments"*. This gives us a single definition which encompasses the common traits in intelligence definitions.

Artificial General Intelligence, sometimes referred as "strong ai", according to Goertzel and Pennachin (2007) means *"AI systems that possess a reasonable degree of self-understanding and autonomous self-control, and have the ability to solve a variety of complex problems in a variety of contexts, and to learn to solve new problems that they didn't know about at the time of their creation."*. It can be seen that an agent fulfilling this definition also possesses the intelligence defined in the previous chapter. In this thesis terms artificial general intelligence and human-level artificial intelligence are treated as synonyms, as they pursue more or less the same goal of general intelligence. Goertzel and Pennachin (2007) suggest that the term AGI is more fitting to the area than HLAI as it human-like approaches are not necessarily used.

The reason general intelligence is specified instead of plain intelligence is that there is a need to differentiate it from the domain specific artificial intelligence, also known as "narrow AI" or "weak AI", that has become prevalent in AI research in recent past. Terms strong AI and weak AI were coined by John Searle in 1980 (John et al. 1980). Narrow AI means smart solutions that may learn and improve their performance through training, but they are only focused on specific type of problems in a specific context. Examples of such AI include chess engines, autonomous vehicles, and natural language processing. These solutions may outperform human capabilities, but only in their limited tasks. When presented a problem outside their domain, they usually perform poorly. As the above definition by Goertzel and Pennachin describes, AGI is able to function on different context and tasks without separate human intervention and reconfiguration.

TODO: check strong ai definition by searle

As The AGI community is diverse and there are multitude of opinions on the best approaches and the goals that should be pursued in the research, several possible roadmaps have been presented in an attempt to create a common basis for the discussion and research of human-level artificial general intelligence. In (Adams et al. March 2012) a high level roadmap with AGI's initial required capabilities and scenario-based milestones is suggested, building on previous work and workshops organized in 2008 and 2009. Presented scenarios can be used to measure the progress and capabilities of AGI restricting the progress of different approaches to a single test situation (Adams et al. March 2012). More concrete example is provided by Ben Goertzel and Gino Yu, who outline creation of a AGI-oriented cognitive architecture based on existing CogPrime architecture (Goertzel and Yu 2014). A simultaneous development of multiple AGI-style applications is suggested to maintain the generality of intelligence. CogPrime is implemented with OpenCog framework, developed by OpenCog Foundation and AI researcher Ben Goertzel. OpenCog is an attempt to create an open source framework for artificial general intelligence (*The Open Cognition Project*; B. Goertzel 2012).

One motivation behind this thesis is to find out if these roadmaps and their suggestions have actually had any effect in the direction the AGI research has gone, or have they been ineffective in their attempt.

### 3 Systematic literature mapping process

Systematic literature mapping is a secondary study method that helps to identify the focal points and research gaps in the subject area, providing an overview of previous research (Petersen et al. 2008). It is also known as systematic mapping study or scoping study (Kitchenham and Charters 2007). This chapter introduces the mapping method, describing each phase of the mapping process. The key differences with a more common study method, systematic literature review (SLR) are also presented, as well as the reasoning behind this method choice.

#### 3.1 Research method description

Systematic mapping is a common research method used in fields such as evidence-based medicine, but have until recently been rare in software engineering (Petersen et al. 2008). Kitchenham, Dybå, and Jorgensen (2004) suggested that adopting evidence-based approach to software engineering research might benefit the industry by ensuring approaches used are backed by evidence. Aggregating evidence is done by systematic literature reviews and similar secondary studies, such as mapping studies (Kitchenham et al. 2010). As researcher bias is one the weak points of secondary studies, adhering to strict protocol and guidelines is required to minimize it (Brereton et al. 2007).

The systematic literature mapping in this thesis is following the method guidelines presented by Petersen et al (2008), later updated in by Petersen, Vakkalanka and Kuzniarz (2015). The mapping process overview can be seen in figure 1. It consists of five separate phases: **definition of research questions, conducting search, screening of papers, keywording, and data extraction and mapping.** Each phase produces a subresult to be used in the next one. This process results in a systematic map of the area. This can and should be further visualized using for example bubble graphs, as it is a powerful way to achieve a quick overview of the field. (Petersen et al. 2008). This also enables easier recognition of research gaps and focus points in the target area.

The process begins by defining focused research questions that are aligned with the goal



Figure 1. Process model (Petersen et al. 2008)

of the study. The goal of the study often is to create a general overview of the research area, and to identify the type and quantity of research (Petersen et al. 2008). Unlike in more focused systematic literature reviews, the research questions of mapping studies are less focused and cover a broader scope (Kitchenham et al. 2010). For example, possible research questions on studies could be: *"Which are the most investigated quality aspects of the software requirements specifications techniques?"* (Condori-Fernandez et al. 2009) or *"What efforts have been reported in literature for Software Engineering curriculum design, implementation, assessment and management?"* (Qadir and Usman 2011). These topic-oriented questions are often combined with research questions regarding the meta-level information, such as publication year, venue, and research methods (Petersen, Vakkalanka, and Kuzniarz 2015).

The next phase of the mapping is the initial material search, which can be conducted multiple ways. Search strings can be formulated from the research questions, and used on academic databases and search engines (Petersen et al. 2008). For example, databases such as IEEE Explore and ACM, as well as aggregators like Google Scholar can be utilized. As the search phrases should be research question driven, following a criteria such as PICO might prove to be helpful, as suggested by Kitchenham and Charters (2007). PICO (Population, Intervention, Comparison, Outcomes) provides an frame to consider research questions' elements and identify keywords. As the goal is to achieve broad overview of the research, study outcomes are not taken into account, as this could result in a biased results (Petersen et al. 2008). The search can also be conducted manually on specific journals and conference proceedings that cover the target area (Petersen et al. 2008). This approach is used in this thesis, as it enables targeting specific reputable and well-known publication venues.

After the initial material has been gathered, it is further refined by excluding papers not rel-

evant to answering the research questions (Petersen et al. 2008). Separate criteria for both inclusion and exclusion is used to find the papers fit for further analysis. According to Petersen, Vakkalanka and Kuzniarz (2015), the criteria may refer to relevance to the topic, publication venue, time period, language restrictions, and evaluation requirements. Considering evaluation requirements should be avoided with systematic maps as it might limit recent trends. Once the criteria is decided, it is applied to the titles and abstracts of the articles (Petersen, Vakkalanka, and Kuzniarz 2015). In case of unclear or poor quality abstracts also the introduction and conclusion sections of the article may be studied (Petersen et al. 2008). As no full-text reading is required, screening of papers can be performed rapidly.

Once the final set of papers is narrowed down and determined, keywording is performed. As described by Peterson (2008), The keywording process starts by first by analyzing the research papers' abstracts by searching possible frequent keywords and prevalent concepts from them. The keywords of each paper are then combined together to achieve a more general set of concepts. On some cases having a more detailed inspection of the article might be required (Petersen et al. 2008; Petersen, Vakkalanka, and Kuzniarz 2015). After the final set of keywords is chosen, they are clustered into categories representing the article population (Petersen et al. 2008). This emergent classification schema can then be used in the data extraction phase.

Different research facets can be used in the classification. For example in addition to the topical scheme emerging from the keywords, a topic-independent facet reflecting the research approach can be used. One example of the latter is the classification of research approaches by Wieringa et. al (2006). Wieringa's classification categorizes scientific papers into six categories such as validation research, solution proposals, and opinion papers. Using existing topic-independent categorization also enables the comparison of different research fields (Petersen, Vakkalanka, and Kuzniarz 2015).

In the final phase of the mapping process, the data extraction is performed by sorting the papers into the classification schemes present (Petersen et al. 2008). The schema may evolve during the data extraction process, changing the categories to match the article population more accurately (Petersen, Vakkalanka, and Kuzniarz 2015). The categorization based on the chosen facets results in a frequency table, i.e. the mapping, with can then be presented

with via visualization and summary statistics. Visualization using for example bubble plots is preferred, as it is a powerful way to represent the information and map of the field (Petersen et al. 2008).

### **3.2 Differences with systematic literature reviews**

Because the systematic literature mapping as a study methods is less known in the field of software engineering, the typical differences between it and more common secondary study method, systematic literature review, is presented here. There is many common factors in both of the methods, and as Kitchenham et al. (Kitchenham et al. 2010) state, *"the distinction between mapping studies and conventional SLRs can be somewhat fuzzy"*. They also present a view that mapping studies are just a different type of systematic literature review.

Similar to other secondary study methods, mapping study aims to summarize and present the research performed in the past. Whereas systematic review focuses on very narrow research questions, mapping study usually has multiple broader questions(Kitchenham et al. 2010). Difference in scope breadth can also be seen in the search strings, as the initial material search for mapping is likely to return large number of studies (Kitchenham and Charters 2007; Petersen et al. 2008). Mappings are usually conducted to achieve an overview of the research area, and therefore the depth of the studies is not as great as in the SLRs. The mapping focuses more on the thematic analysis of the articles instead of in-depth analysis of their results or gathering empirical evidence based on their results, which is often the goal of an traditional SLR (Petersen et al. 2008). Therefore the quality of the objects of study is not relevant, unless the quality itself is the aspect to be investigated.

Since both methods have their strengths and weaknesses, using them complementarily can be an effective combination. As suggested by Petersen et. al (2008) and Kitchenham et al. (Kitchenham et al. 2010), good approach is to first get an overview of the research area with systematic map, and then applying conventional literature review to a specific focus area. Results of the mapping can provide information on the quantity of available evidence, which can help targeting the follow-up SLRs more precisely.

### **3.3 Background/reasoning behind method choice**

As a thesis topic, Artificial General Intelligence is a challenging and broad area. By focusing on the AGI research articles as an object of study, useful research can still be performed without requiring too much prior knowledge and expertise on the topic from the author. The reason behind the choice of systematic literature mapping is that it fits specifically well on creating overview of the study area. As the research questions of this thesis are broad, mapping study is a more suitable approach than a conventional SLR.

Due to the complexity of the topic, it can be difficult to enter as a newcomer. Kitchenham et al. (2010) suggest that a systematic mapping of the field can be useful to researchers new to the area. It can also be useful in introducing it to people unfamiliar with the field, both in academia and on the industry's side. As can be seen from the history of AI research in 2.1, AGI is also a topic that is known of its fluctuating interest and popularity, therefore seeing the temporal trends and state of the current research can be of interest. Mapping study is a good way to achieve that.

The mapping process guidelines by Petersen et al. (2008; 2015) were chosen to be followed as they were the most used ones in software engineering community (Petersen, Vakkalanka, and Kuzniarz 2015). Also the most popular guideline for systematic literature review by Kitchenham et al. (2007) can be utilized to some degree because of the similarity of the methods. These guidelines provide clear methodologies for the study. In addition, the process can further utilize for example the scientific paper classification scheme presented by Wieringa et al. (Wieringa et al. 2006), as it provides an useful and straightforward research facet to the categorization phase of the process.



## 4 Conducting the literature mapping

**The following text is very raw, will be edited properly later**

In this chapter the systematic literature mapping is conducted. To ensure that the study is conducted with rigor, documenting it in detail is essential. This chapter only describes the process, the actual results of the mapping are presented in chapter 5. Following the guidelines mentioned in section 3.1, the following sections focus on each of the phases of the study.

A search for similar literature studies was performed on University of Jyväskylä's thesis archive JYX, Google Scholar, and Scopus, and no results were found. Studies that were found were mostly conventional SLRs with much more narrow scope. It was concluded that there exists no similar mapping study on the topic.

TODO: mention each subsection in more detail?

### 4.1 Research questions

The following research questions aim to cover the goal of the study, to create an overview of the scientific research performed on the field of AGI.

RQ1: How much, and what kind of research is done in the field of AGI?

RQ2: Where and when were the studies published?

RQ3: What are the major research topics in the field and have they changed over time?

The first question focuses on the type and volume of the published AGI related research. The second question focuses on the publication venue and time of the articles, showing temporal trends and popular forums for the topic. Research question three aims to find out the most popular subtopics and their change in the study's time frame.

TODO: how these are answered, does it need to be addressed here or only later?

## 4.2 The material search

This thesis utilized both manual and broad automatic search in the material gathering. The targeted publication venues are either very topic specific, or otherwise high-ranked in the artificial intelligence community, and general content-wise. The following venues are used in this thesis:

- **Journal of Artificial general intelligence** (journal)
- **Artificial Intelligence** (journal)
- **Journal of Artificial Intelligence Research** (journal)
- **International Conference on Artificial General Intelligence** (conference)
- **International Joint Conference on Artificial Intelligence** (conference)

One basis for the selection of publications was the Finnish Publication Forum (*Publication Forum*) ranking. In this thesis Publication Forum is referred as JUFO. Publication Forum ranks peer-reviewed scientific publication venues on grades 1-3, with 1 being the standard level, and 3 being the highest.

The reasoning behind venue choices were as follows: Journal of Artificial general intelligence (JAGI) (JUFO rank 1) and International Conference on Artificial General Intelligence (ICAGI) are very topic specific, and therefore should contain the most relevant information within the field. These main forums are likely to contain most of the research articles. Artificial Intelligence journal (AIJ) and Journal of Artificial Intelligence Research (JAIR) were chosen because they have the highest JUFO ranking and therefore are leading publications of the AI field. Both also have high citescore (7.7 and 6.4 respectively) considering they do not focus on any specific subfield of AI. For the same reason, International Joint Conference on Artificial Intelligence (IJCAI) was included (JUFO rank 2). These more general sources may not contain many papers as the topic-specific publications, but can show the relative popularity of the field within the AI research.

TODO: more basis?

Out of these, Journal of Artificial General Intelligence (JAGI) was searched manually due to low quantity of articles, but otherwise automatic search was performed using search engines

Scopus and Springer Link.

The publication interval of the papers was limited to years **2015-2019** to reduce the material to reasonable quantity. This choice prevents seeing temporal trends that might not show in such a short interval, but longer time frame was not viable considering this is only a master's thesis.

#### 4.2.1 Search phrases

The search phrases for the automatic search were derived from the research questions, following the aim of this study. Included were the close concepts similar to AGI, such as HLAI and superintelligence. The logical search terms used are listed on table 1. Search term 1 is very broad and non-restricting, but this helps it yield results even in more general venues.

#	Logical search term
#1	"artificial general intelligence" OR agi OR "human-level ai" OR HLAI OR superintelligence
#TODO	If no more than one row is needed, maybe remove this table?

Table 1: Search phrases

- evaluation of search: test set of known papers, possible here?

### 4.3 Choosing papers for inclusion

The initial material search was performed during October 2020. The papers found on the search were further narrowed down using the following criteria. All literature accepted for this thesis should be:

- presenting and/or relating to Artificial General Intelligence, Human-level AI, or similar concept of general intelligence,
- written in English,
- published between 2015-2019,
- a peer-reviewed journal or conference paper, and

- available digitally free or as a member of University of Jyväskylä.

Piece of literature should be excluded if it is:

- a book, a thesis, or lecture notes,
- not wholly available digitally or
- a duplicate of already included paper.

For a paper to be found relevant to the topic, it needed to clearly reference the concept of (artificial) general intelligence in its title or abstract, with intention of relating the article to it. If the concept is only mentioned as a sidenote or as an example without further relevance, it should be excluded. If a clear decision couldn't be made from these parts, article could be further examined. In case of duplicate articles or articles concerning the same study, only the latest one is included in the study.

The initial search yielded total of 187 papers. Almost all of the initial results fulfilled the meta-level criteria as the used search engines (Scopus and Springer Link) offered extensive filtering capabilities. These results were then inspected in two separate phases. In the first phase, the titles and abstracts of the articles were examined, and then the article was marked as either excluded or potential in a spreadsheet. After this initial inspection, 122 papers were considered potential, and would be subject of further inspection.

In the second phase, the potential papers were examined again, with the research questions and inclusion criteria considered more thoroughly, to further exclude papers that were not relevant to the topic. During this phase, a small example set of excluded and included papers was sent to the thesis councilor to confirm their categorization. During this phase it became apparent that without stricter inclusion criteria the amount of accepted papers would be excessively high. It was decided that to be included, the article would have to explicitly relate itself to general intelligence. In the end, 92 papers were included and would be used as the mapping sample. Table 2 shows the number of accepted papers and their publication venues.

Publication forum	Total results	Potential results	Accepted results	Accepted results (%)
JAIR	5	3	2	40%
IJCAI	12	8	4	33.33%
AIJ	3	3	1	33.33%
JAGI	15	9	6	40%
ICAGI	152	99	79	51.97%
<b>All</b>	<b>187</b>	<b>122</b>	<b>92</b>	<b>49.20%</b>

Table 2: Publication forums and their results

#### 4.4 Keywording

Once the articles were narrowed down to those relevant to topic, they were inspected more thoroughly and the emergent concepts or keywords were extracted. These keywords were then further combined to topic categories until no further merging could be done without making them too general and losing too much information in the process. The resulting categorical scheme of 22 distinct topic categories is illustrated in Table 4 in the results chapter.

#### 4.5 Data extraction and mapping

In the actual data extraction phase, the included research articles were sorted into two different categorization schemes. First one was the topic categorization that emerged from the papers themselves, and the other one was the topic-independent categorization presented by Wieringa et al. (2006). The latter enables us to make comparisons with AGI and other fields of research that aren't necessarily related. It also provides information on the kind of research done, answering RQ1. Wieringa's scheme is presented in Table 3.

<b>Class</b>	<b>Class description</b>
Evaluation research	Investigates existing problem or technique in practice with sound research methods.
Validation research	Investigates a solution proposal not yet implemented in practice using sound research methods. Prototypes, simulations or mathematical proof are among possible methods.
Solution proposal	Paper proposes a new solution to a problem, with arguments for its relevance. Solutions must be novel or significantly improve existing ones.
Philosophical papers	Conceptual frameworks and new viewpoints on the topic.
Opinion papers	Papers presenting author's personal opinion on the topic.
Personal experience papers	Present author's experience on something that has been done in practice. Doesn't necessarily provide evidence.

Table 3: Article classification by Wieringa et al. (2006)

The papers could belong to multiple categories on both schemes. Although this decision makes some graphs slightly less accurate(?), provides it information on which topics and what kinds of papers are usually seen together. The results of this mapping process are presented in the following chapter.

## 4.6 Source material control

TODO: Maybe don't make this its own subsection, just a paragraph should be fine.

- How the papers were handled - How graphs etc. were made - Other meta level information
- Excel, python, etc.

## **5 Results and analysis**

In this chapter the results of the literature mapping are presented. Each of the research questions is answered with the corresponding results, and observations made during the study are shown. TODO: SOMething better here

### **5.1 Publications years and venues**

As the field of Artificial General Intelligence is relatively small when compared to other branches of AI research, it is very focused on specific publication forums. This can be seen evidently from the sample papers' publication venues. Figure 2 shows that almost 86% of the 92 sample articles were published as the proceedings of the International Conference on Artificial General Intelligence. The second largest concentration of articles was on the Journal of Artificial General Intelligence, as was expected it being the only non-conference publication on the topic. It needs to be also noted that even though Artificial Intelligence Journal and Journal of Artificial Intelligence Research are esteemed journals that contribute hundreds of AI articles every year, their articles only constituted less than four percent of this study's papers.

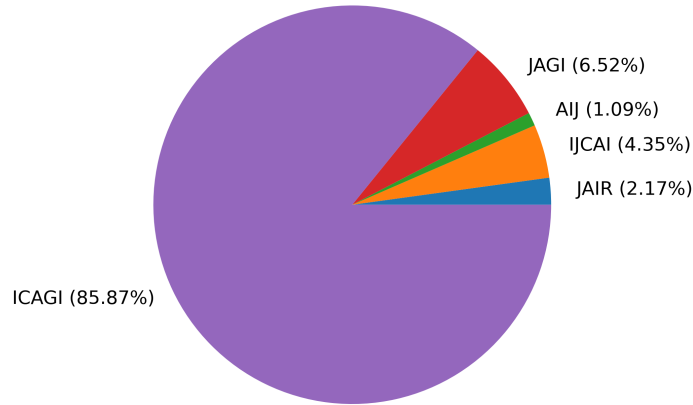


Figure 2. Article distribution between the publication forums

In Figure 3 the amount of yearly published articles and their venues are visualized in a bar plot. As the inspected time period on this thesis is short, only five years, predicting temporal trends accurately is not possible. However, it can be seen that every year there is steady publication pace of at least 13 articles, with the average being 18.4 a year. In year 2018 25 different articles were published, which was the highest number during the inspection period.

ICAGI is the *"only major conference series devoted wholly and specifically to the creation of AI systems possessing general intelligence at the human level and ultimately beyond"* (AGI Conference). It has been organized yearly since 2008.

Therefore to answer research question RQ2, it can be said that while there are AGI studies being published in different publication forums, vast majority of the research is released as yearly conference proceedings of International Conference on Artificial General Intelligence. Research community is active but small, as only average of 18,4 articles fitting this scope are published yearly.



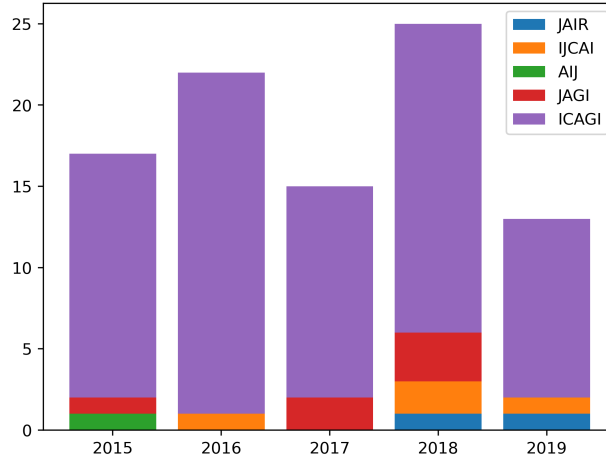


Figure 3. Yearly articles by publication forum

## 5.2 Common research topics

The main goal of this thesis was to achieve an overview of the AGI field. Through the mapping process, 22 different topic categories were found, and they are presented in Table 4. A single paper can relate to multiple topics. These topics show us how the AGI research is focused on the top level. From the amount of different topics it can be seen that the research area is broad and not focused on a single style of solutions.

From the sample articles, four different themes could be identified. These themes are just high-level observations by the author based on the found topic categories. **Building AGI systems** is of course naturally a very prominent one. These papers were more focused on the possible theory and implementations that could be used in building of an AGI system: how it reasons and plans, how capabilities like perception could be implemented, and how to handle issues like combinatorial explosion. One reoccurring theme was **Learning**. Different approaches to more common machine learning techniques like reinforcement learning and algorithms related to it like Q-learning and Deep Q-Networks are demonstrated. In additions to this, research on cumulative learning, lifelong learning, and episodic memory is very prominent. Third theme was **Agent interaction**, meaning topics such as multi-agent systems, environmental interaction, and human-computer interaction. Lastly there are **Non-**

**technical topics**, such as philosophical questions, AI ethics, emotion in AI and meta-level research.

#	Category	Category description
1	Cognitive architectures	Cognitive architectures and their descriptions.
2	AGI Design	General ideas on how AGI or its components could be designed and implemented.
3	Reasoning and Inference	Approaches on temporal and causal reasoning and inference techniques.
4	Planning and decision making	Utilizing existing knowledge in planning and decisions.
5	Probabilistic approaches	Probabilistic approaches e.g. Bayesian techniques and uncertainty handling.
6	Category theory	Approaches relating to category theory.
7	Universal AI	Concepts relating to Universal AI (Hutter 2004): universal induction, AIXI, compression.
8	Physical robots	Physical robots and interaction with physical environment.
9	Computer vision and perception	Topics concerning vision and perception systems of an agent.
10	Nature-inspired approaches	Artificial animals, homeostatic agents and other nature-inspired ideas.
11	Reinforcement learning	Topics directly relating to reinforcement learning, e.g. Q-learning, rewarding techniques.
12	Recursive self-Improvement	Relating to fast self-improvement of an agent and intelligence explosion.
13	Experiential learning	Cumulative learning, artificial pedagogy and other topics related to how agent builds on existing knowledge.
14	Agent environment	Descriptions of environments and how agents interact within them.
15	Multi-agent systems	Topics relating to agent-to-agent interaction and cooperation.
16	Human-computer interaction	How human and agent interact and communicate and their relation to each other.
17	AI safety	Approaches on how to safely create and interact with AGI, and what safety issues arise alongside general intelligence.
18	Philosophical aspects	Philosophical questions relating to artificial intelligence, e.g. AI ethics and morality.
19	Human-like qualities	Approaches with basis on human qualities like emotion and empathy.
20	AGI research	Secondary research about AGI research.
21	AI evaluation	How to evaluate and measure AI intelligence and performance.
22	Game playing	Game playing as a tool in development and evaluation of general agents.

Table 4: Emergent topic categories

In Figure 4, the frequencies of research topics is presented over the years. The most researched topic is **Cognitive architectures**. Cognitive architecture means an abstract model of cognition as well as its software implementation, which aims to be a system showing intelligent behaviour, artificial intelligence (Lieto et al. 2018). In AGI research there are few cognitive architectures that are standing out in the field. Goertzel’s OpenCog framework (See section 2.2) was seen in 8 different papers, e.g Goertzel’s idea of bridging the gap between theory and practice in AGI design (Ben Goertzel 2017) and Potanov’s attempt to create semantic vision system by combining OpenCog with YOLOv2 object detection system (Potapov et al. 2018). In addition to OpenCog, Non-Axiomatic Reasoning System (NARS), developed by Pei Wang, was part of many articles. For example, besides the introduction of its implementation (Hammer, Lofthouse, and Wang 2016), papers describing its approach to emotion (Wang, Talanov, and Hammer 2016) and inferential learning (Wang and Li 2016). While many of these cognitive architecture articles were mainly focused on presenting authors’ system’s implementation, some were offering ideas that could be used in any other approaches to AGI. These articles, among other papers focusing on not-implementation-specific ideas, were categorized also to **AGI design** category that included 6 papers.

The second largest topic category was **Universal AI**. The theory of universal AI was created by Marcus Hutter (2004) and it describes a complete mathematical model for general artificial intelligence, named AIXI (Hutter 2012). Although incomputable, this theory and topics related to it are target of vigorous research. In this category there is 14 articles concerning AIXI, Solomonoff’s universal induction, functional programming and compression. For example in 2019 paper by Franz, Gogulya and Löffler (2019), a monolithic inductive approach to AGI was presented, taking advantage of AIXI and incremental compression techniques. In (Martin, Everitt, and Hutter 2016), death was formally defined to generally intelligent agents like AIXI and discoveries were made regarding agents’ behaviour in such situations.

Use susumu’s paper to link aixi to rl, fp etc. MAgichaskeller?

CA, universal ai, reinforcement learning, experiential learning, ai safety, philosophical asps

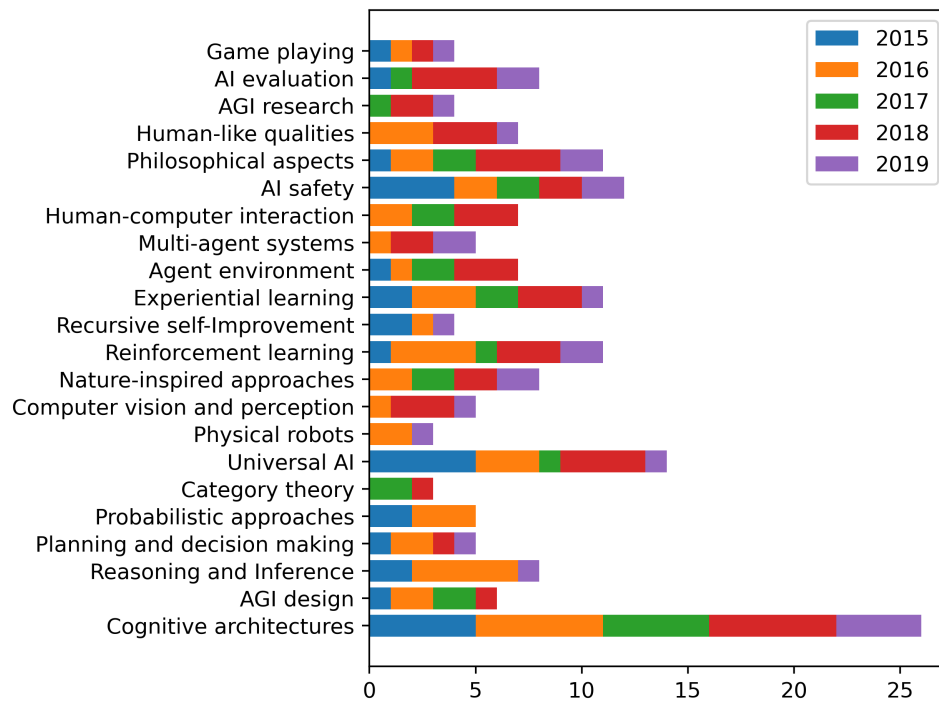


Figure 4. Topic category frequencies

Here's something about the short time trends that can be seen from graphs.

Here more text about the topics, and a heatmap about the topic categories.

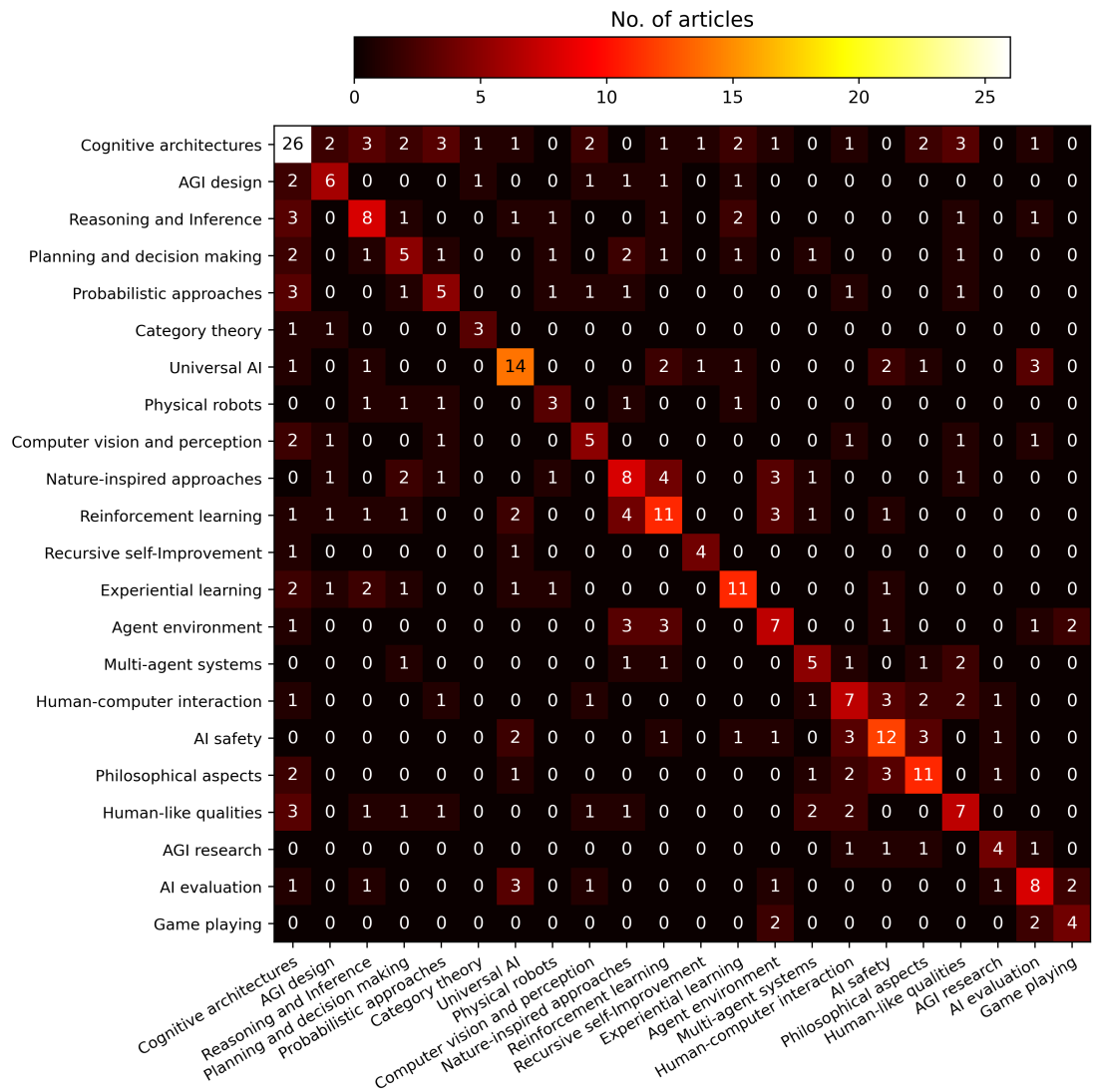


Figure 5. Topic's relations to each other

### 5.3 Types of AGI research

Here is text telling about the bubble plot and how the classes and topics relate to each other.



Figure 6. Bubble plot showing article distribution between topics and article types

## **5.4 TODO: Where to put this: Validity threats?**

- performed alone, researcher bias etc (multiple people not possible here), how rigor is achieved, repeatability,

Good source on reliability Wohlin et al. 2013

## **5.5 TODO: Where to put this: Possible continuation research**

- List of most prominent topics for further research

## **6 Conclusion**

In this thesis, a systematic literature mapping was conducted on the field of artificial general intelligence. Results of the study showed that ....



## Bibliography

Adams, Sam, Itmar Arel, Joscha Bach, Robert Coop, Rod Furlan, Ben Goertzel, J. Storrs Hall, et al. March 2012. "Mapping the Landscape of Human-Level Artificial General Intelligence". *AI Magazine* 33, number 1 (): 25–42. <https://doi.org/10.1609/aimag.v33i1.2322>. <https://www.aaai.org/ojs/index.php/aimagazine/article/view/2322>.

*AGI Conference*. Accessed: 2021-01-14. <https://www.agi-conference.org/>.

Anderson, M. 2006. "Intelligence". *MS Encarta online encyclopedia*.

Brereton, Pearl, Barbara A Kitchenham, David Budgen, Mark Turner, and Mohamed Khalil. 2007. "Lessons from applying the systematic literature review process within the software engineering domain". *Journal of systems and software* 80 (4): 571–583.

Buchanan, B, E Feigenbaum, and J Lederberg. 1968. "Heuristic DENDRAL-A program for generating explanatory hypotheses in organic chemistry."

Cheeseman, Peter C. 1985. "In Defense of Probability." In *IJCAI*, 2:1002–1009.

Condori-Fernandez, N., M. Daneva, K. Sikkil, R. Wieringa, O. Dieste, and O. Pastor. 2009. "A systematic mapping study on empirical evaluation of software requirements specifications techniques". In *2009 3rd International Symposium on Empirical Software Engineering and Measurement*, 502–505.

Franz, Arthur, Victoria Gogulya, and Michael Löffler. 2019. "WILLIAM: A monolithic approach to AGI". In *International Conference on Artificial General Intelligence*, 44–58. Springer.

Goertzel, B. 2012. "CogPrime : An Integrative Architecture for Embodied Artificial General Intelligence".

Goertzel, Ben. 2017. "From abstract agents models to real-world AGI architectures: bridging the gap". In *International Conference on Artificial General Intelligence*, 3–12. Springer.

Goertzel, Ben, and Cassio Pennachin. 2007. *Artificial general intelligence*. Volume 2. Springer.

- Goertzel, Ben, and Gino Yu. 2014. "From here to AGI: A roadmap to the realization of human-level artificial general intelligence". *2014 International Joint Conference on Neural Networks (IJCNN)*, 1525–1533.
- Haenlein, Michael, and Andreas Kaplan. July 2019. "A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence:" [inlangen]. *California Management Review* (). <https://doi.org/10.1177/0008125619864925>. [https://journals-sagepub-com.ezproxy.jyu.fi/doi/full/10.1177/0008125619864925?casa\\_token=cJ1YEP0BQQQAAAAA%3ANQqrLNDBpfNEU7WmxeiyQ178cT8ONGu70KePNj4i-XuylxQTKlbnFz11d1rXHTtf9kNb\\_GKVXmnZrw](https://journals-sagepub-com.ezproxy.jyu.fi/doi/full/10.1177/0008125619864925?casa_token=cJ1YEP0BQQQAAAAA%3ANQqrLNDBpfNEU7WmxeiyQ178cT8ONGu70KePNj4i-XuylxQTKlbnFz11d1rXHTtf9kNb_GKVXmnZrw).
- Hammer, Patrick, Tony Lofthouse, and Pei Wang. 2016. "The OpenNARS implementation of the non-axiomatic reasoning system". In *International conference on artificial general intelligence*, 160–170. Springer.
- Hutter, Marcus. 2004. *Universal artificial intelligence: Sequential decisions based on algorithmic probability*. Springer Science & Business Media.
- . 2012. "One decade of universal artificial intelligence". In *Theoretical foundations of artificial general intelligence*, 67–88. Springer.
- John, Searle, et al. 1980. "Minds, Brains and Programs". *Behavioral and Brain Science* 3:417–424.
- Kitchenham, Barbara, and Stuart Charters. 2007. *Guidelines for performing systematic literature reviews in software engineering*. Technical report.
- Kitchenham, Barbara, Rialette Pretorius, David Budgen, O Pearl Brereton, Mark Turner, Mahmood Niazi, and Stephen Linkman. 2010. "Systematic literature reviews in software engineering—a tertiary study". *Information and software technology* 52 (8): 792–805.
- Kitchenham, Barbara A, Tore Dyba, and Magne Jorgensen. 2004. "Evidence-based software engineering". In *Proceedings. 26th International Conference on Software Engineering*, 273–281. IEEE.
- Legg, Shane, Marcus Hutter, et al. 2007. "A collection of definitions of intelligence". *Frontiers in Artificial Intelligence and applications* 157:17.

- Lieto, Antonio, Mehul Bhatt, Alessandro Oltramari, and David Vernon. 2018. *The role of cognitive architectures in general artificial intelligence*.
- Lighthill, James. 1973. "Artificial intelligence: A general survey". In *Artificial Intelligence: a paper symposium*, 1–21. Science Research Council London.
- Liu, Jiaying, Xiangjie Kong, Feng Xia, Xiaomei Bai, Lei Wang, Qing Qing, and Ivan Lee. 2018. "Artificial intelligence in the 21st century". *IEEE Access* 6:34403–34421.
- Martin, Jarryd, Tom Everitt, and Marcus Hutter. 2016. "Death and suicide in universal artificial intelligence". In *International Conference on Artificial General Intelligence*, 23–32. Springer.
- McCarthy, John. 2007. "From here to human-level AI". *Artificial Intelligence* 171 (18): 1174–1182.
- McCulloch, Warren S, and Walter Pitts. 1943. "A logical calculus of the ideas immanent in nervous activity". *The bulletin of mathematical biophysics* 5 (4): 115–133.
- Minsky, Marvin. 2007. *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind*. Simon / Schuster.
- Minsky, Marvin L, Push Singh, and Aaron Sloman. 2004. "The St. Thomas common sense symposium: designing architectures for human-level intelligence". *Ai Magazine* 25 (2): 113–113.
- Myers, Ware. 1986. "Introduction to expert systems". *IEEE Computer Architecture Letters* 1 (01): 100–109.
- Newell, Allen, John C Shaw, and Herbert A Simon. 1959. "Report on a general problem solving program". In *IFIP congress*, 256:64. Pittsburgh, PA.
- Nilsson, Nils J. 2005. "Human-level artificial intelligence? Be serious!" *AI magazine* 26 (4): 68–68.
- . 2009. *The quest for artificial intelligence*. Cambridge University Press.
- Pearl, Judea. 1988. "Probabilistic reasoning in intelligent systems: Networks of plausible inference". *Morgan Kauffmann Publishers Inc*.

Petersen, Kai, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. 2008. "Systematic mapping studies in software engineering". In *12th International Conference on Evaluation and Assessment in Software Engineering (EASE) 12*, 1–10.

Petersen, Kai, Sairam Vakkalanka, and Ludwik Kuzniarz. 2015. "Guidelines for conducting systematic mapping studies in software engineering: An update". *Information and Software Technology* 64:1–18.

Potapov, Alexey, Innokentii Zhdanov, Oleg Scherbakov, Nikolai Skorobogatko, Hugo Lapie, and Enzo Fenoglio. 2018. "Semantic image retrieval by uniting deep neural networks and cognitive architectures". In *International Conference on Artificial General Intelligence*, 196–206. Springer.

*Publication Forum*. Accessed: 2020-10-20. Federation of Finnish Learned Societies. <https://www.julkaisufoorumi.fi/en>.

Qadir, M. M., and M. Usman. 2011. "Software Engineering Curriculum: A systematic mapping study". In *2011 Malaysian Conference in Software Engineering*, 269–274.

Rumelhart, David E, Richard Durbin, Richard Golden, and Yves Chauvin. 1995. "Backpropagation: The basic theory". *Backpropagation: Theory, architectures and applications*, 1–34.

Russell, Stuart, and Peter Norvig. 2009. *Artificial Intelligence: A Modern Approach*. 3rd. Prentice Hall Press. ISBN: 0136042597.

Shortliffe, Edward H, Randall Davis, Stanton G Axline, Bruce G Buchanan, C Cordell Green, and Stanley N Cohen. 1975. "Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system". *Computers and biomedical research* 8 (4): 303–320.

Silver, David, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. 2016. "Mastering the game of Go with deep neural networks and tree search". *nature* 529 (7587): 484–489.

Silver, David, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dhharshan Kumaran, Thore Graepel, et al. 2018. "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play". *Science* 362 (6419): 1140–1144.

*The Open Cognition Project*. Accessed: 2020-09-24. [https://wiki.opencog.org/w/The\\_Open\\_Cognition\\_Project](https://wiki.opencog.org/w/The_Open_Cognition_Project).

TURING, A.M. 1950. "Computing machinery and intelligence". *Mind* 59 (236): 433–460.

Wang, Pei. 1995. "On the working definition of intelligence". *Center for Research on Concepts and Cognition CRCC, Indiana University*.

Wang, Pei, and Xiang Li. 2016. "Different conceptions of learning: function approximation vs. self-organization". In *International Conference on Artificial General Intelligence*, 140–149. Springer.

Wang, Pei, Max Talanov, and Patrick Hammer. 2016. "The emotional mechanisms in NARS". In *International Conference on Artificial General Intelligence*, 150–159. Springer.

Widrow, Bernard, and Michael A Lehr. 1995. "Perceptrons, Adalines, and backpropagation". *Arbib* 4:719–724.

Wieringa, Roel, Neil Maiden, Nancy Mead, and Colette Rolland. 2006. "Requirements engineering paper classification and evaluation criteria: a proposal and a discussion". *Requirements engineering* 11 (1): 102–107.

Wohlin, Claes, Per Runeson, Paulo Anselmo da Mota Silveira Neto, Emelie Engström, Ivan do Carmo Machado, and Eduardo Santana de Almeida. 2013. "On the reliability of mapping studies in software engineering". *Journal of Systems and Software* 86 (10): 2594–2610. ISSN: 0164-1212. <https://doi.org/https://doi.org/10.1016/j.jss.2013.04.076>. <http://www.sciencedirect.com/science/article/pii/S0164121213001234>.