



HAL
open science

An Initial Generic Assessment Framework for the Consideration of Risk in the Implementation of Autonomous Systems

K. Tara Smith, Lynne Coventry, Robert Greensmith

► **To cite this version:**

K. Tara Smith, Lynne Coventry, Robert Greensmith. An Initial Generic Assessment Framework for the Consideration of Risk in the Implementation of Autonomous Systems. 5th IFIP Working Conference on Human Work Interaction Design (HWID), Aug 2018, Espoo, Finland. pp.186-199, 10.1007/978-3-030-05297-3_13. hal-02264622

HAL Id: hal-02264622

<https://inria.hal.science/hal-02264622v1>

Submitted on 7 Aug 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

An Initial Generic Assessment Framework for the Consideration of Risk in the Implementation of Autonomous Systems

K Tara Smith¹, Dr Lynne Coventry² and Robert GreenSmith¹

¹ Human Factors Engineering Solutions Ltd, PO Box 29175, Dunfermline, KY12 9XU, Scotland

² Northumbria University, Newcastle Upon Tyne, NE1 8ST

lncs@springer.com

Abstract. This paper considers some of the issues around autonomous systems and the different types of risk involved in their implementation. These risks are both barriers to the implementation of a successful autonomous system and risks that are consequences of the use of such systems. The different levels of automation, and different approaches to categorizing these levels, as presented in a variety of frameworks, are summarized and discussed.

The paper presents an initial generic assessment structure, with the aim of providing a useful construct for the design and development of acceptable autonomous systems that are intended to replace elements of the human cognitive process, specifically in situations involving decision-making. It introduces the concept of the “logos chasm”: the gap between achievable autonomous systems and those which currently only exist in the realm of science fiction; and discusses possible reasons for its existence.

Keywords: Autonomous Systems, Automation Risks, Automation Frameworks.

1 Introduction

There has been a considerable amount of work looking at defining a useful framework or model for the consideration of autonomous systems. Much of this work has been domain specific, such as those being applied to cars or Unmanned Aerial Vehicle Systems (UAS). These frameworks are generally concerned with the extent of the automation within the system: i.e. how much of the task is performed by the automation and how much by a human operator. However, when actually implementing an autonomous system, there is another critical concern: the capability of the technology.

This paper presents an initial structure for combining these two aspects, allowing an assessment of whether the current state of automation technology will support the envisaged level of autonomy, together with an initial consideration of the level of risk associated with implementing that level of automation in that circumstance. This

structure is intended to be applicable across any domain and to function as an early contribution to a design or development project; effectively grounding it in reality rather than science fiction. The structure also allows for the potential for technological advances, providing some degree of “future-proofing”.

The proposed assessment structure is used as a mechanism for discussing some of the wider issues around autonomous systems, considering the impactors on their implementation and use and some possible reasons for risks and limitations. It concentrates on systems with a fixed level of autonomy rather than those implementing dynamic adjustable autonomy (where the human is in control of the level of autonomy at any one point) or adaptable autonomy (when the computer decides upon the appropriate level of autonomy), however at the point of any single decision the system will have a single level of autonomy implemented.

2 Background

System autonomy is the capabilities that enables a particular system to be self-determining, or within designed limits, self-governing. Automation is the mechanism or set of mechanisms that support an autonomous, or self-governing system.

A number of models of autonomy exist, which present different levels of system autonomy ranging from no autonomy, i.e. a completely manual system directed by a human operator to a completely autonomous system, i.e. the human is completely out of the loop. This section provides an overview of a few of these, illustrating the different approaches adopted. An example of a simple model is the 4 layers proposed by Endsley [1].

Table 1. Taxonomy Proposed By Endsley (1987)

| Level of autonomy | Description | Explanation |
|--------------------------|--|---|
| Level 1 | Decision support | Human acts upon system recommendations. |
| Level 2 | Conceptual Artificial Intelligence (AI) | The system acts autonomously however, the consent of the operator is required to carry out actions. |
| Level 3 | Monitored AI | The system acts autonomously unless vetoed by the human. |
| Level 4 | Fully autonomous with no operation interaction | The system excludes the human from the loop-Fully autonomous case |

At the other end of the continuum, versions of 10 levels of automation are presented, for example, in Sheridan's Model of Autonomy [2].

Table 2. Sheridan's Model of Autonomy

| Level of autonomy | Explanation |
|--------------------------|---|
| Level 0 | The computer offers no assistance: human must take all decisions and action |
| Level 1 | The computer acquires the data from the process, and registers them without analysis (1* new level) |
| Level 2 | The computer offers a complete set of decision/action alternatives, human decides and acts |
| Level 3 | The computer narrows the selection down to a few alternatives, human decides and acts |
| Level 4 | The computer suggests one alternative, human acts |
| Level 5 | The computer executes one alternative if the human approves, or |
| Level 6 | The computer allows the human a restricted time to veto before automatic execution, or |
| Level 7 | The computer executes automatically, then necessarily informs the human, |
| Level 8 | The computer executes automatically, informs the human only if asked, or |
| Level 9 | The computer executes automatically, informs the human only if it, the computer, decides to |
| Level 10 | The computer decides everything, acts autonomously, ignoring the human |

A different approach, such as that presented by Sheridan [2] and Parasuraman [3] is to describe the extent of the autonomy in terms of the level of supervisory control provided by a human operator, which is summarised in Table 3 below.

Table 3. Summary Levels of Supervisory Control for Adjustable Autonomy

| Level | Description |
|-----------------------|--|
| Manual Control | The UAS is responsible for gathering and displaying unfiltered, unprioritized information for the human. The human still is the prime monitor for all information, responsible for filtering, prioritizing, and assessing the data). |

| Level | Description |
|--|---|
| Consent Based Autonomy | The UAS gathers, filters, and prioritizes information displayed to the human, in time for human--operator to provide consent or to intervene. |
| Executive Autonomous Operations | UAS observes and monitors all systems and commands and acts autonomously, informing the human operator after the fact, displaying information only if asked, ignoring the human |

Other models consider autonomy at the level of functions within a system, such as NASA's Function Specific Level of Autonomy Tool (FLOAT) [4], which also categorises the levels against the OODA (Observe, Orient, Decide, Act) loop [5].

Table 4. NASA's 5-pt Scale and Definition of the Levels of Autonomy (from Proud and Hart 2005)

| Level | Observe | Orient | Decide | Act |
|--------------|--|--|--|--|
| 1 | The data is monitored on the ground without assistance from onboard | The calculations are performed on the ground without assistance from onboard | The decision is made on the ground without assistance from onboard | The task is executed by ground support without assistance from onboard |
| 2 | The majority of the monitoring will be performed by ground support with available assistance onboard | The majority of the calculations will be performed by ground support with available assistance onboard | The decision will be made by ground support with available assistance onboard | The task is executed by ground support with available assistance onboard |
| 3 | The data is monitored both onboard and on the ground. | The calculations are performed both onboard and on the ground. | The decision is made both onboard and on the ground and the final decision is negotiated between them. | The task is executed with both onboard and ground support |
| 4 | The majority of the monitoring will be performed onboard with available assistance from ground support | The majority of the calculations will be performed onboard with available assistance from ground support | The decision will be performed onboard with available assistance from ground support | |

| Level | Observe | Orient | Decide | Act |
|-------|--|---|---|---|
| 5 | The data is monitored onboard without assistance from ground support | The calculations are performed onboard without assistance from ground support | The decision is made onboard without assistance from ground support | The task is executed onboard without assistance from ground support |

These models are just a small selection of those that have been developed to illustrate a flavour of their similarities and differences.

Taking the gamut of the models, they could be classified as covering four main dimensions:

- the level of completeness of the autonomy
- the categorisation of the level of control by the human
- the relationship to the cognitive processes covered
- the relationship to the type of task being undertaken.

2.1 Relevance to Human-Computer Interaction (HCI)

The world today is characterised by increasing automation and digitization. Systems are becoming distributed and interconnected, with invisible and complex processes. Humans working and living in these environments are confronted with new challenges and are sometimes blamed for the failure of these systems. It is important to realise that building reliable and robust complex human-machine or socio-technical systems requires going beyond blaming the human and attempting to replace them with full-blown automation, leaving only a residual role for humans added as an afterthought when complete automation is impossible. Instead, “we need to develop our designs from the outset to take advantage of some of the wonderful flexibilities and capabilities of human beings” [6]. This has led to the development of a more human-centred approach to automation. This approach acknowledges an increasing awareness of the importance of human skills and judgment in making such systems work and the role of the interfaces in ensuring that the human is fully aware of the current situation and able to either take over, or trust in the automation; depending on the level of automation.

As an example, the automobile industry is focusing on pushing technology towards fully automated driving as fast as possible. This inevitably leads to a technology focus, neglecting aspects of the user experience, acceptance and trust. To fully take advantage of automobile interaction, whatever the level, requires further human-computer interaction research to understand how to best support drivers and passengers. Kun et al [7] propose a research agenda for automotive user interfaces (UIs) in the age of automated driving.

The automobile industry is not the only area where artificial intelligence is giving increasing levels of autonomy. Artificial intelligence already has some level of auton-

omy over filtering our spam [8], deciding what we see on social media [9], providing legal advice [10], predicting future crimes [11] and delivering health care.

Health care is the current domain looking to exploit artificial intelligence and facilitate increasing autonomy. While it is hard to find examples of fully autonomous systems delivering treatment to patients, most progress has been made with insulin pumps [12]. However advances have also been made in automating cancer diagnosis (e.g. Platania et al 2017 [13]).

3 Overview of Proposed Autonomy Assessment Structure

The preceding section shows that there are different taxonomies for describing the intent and capability levels of autonomy, but each one tends to approach the question from a single perspective: some of them then go on to assess aspects of risk.

The assessment structure presented here approaches the question from two different perspectives, with the aim of assessing the level of risk to the success of the implementation of an autonomous system in terms of its automation components. The approach is summarised as:

- Description of intent
- Description of capability
- Type of operational context
- What capability is needed for this intent
- What level of risk is implied by the combination of the selected level of intent and the capability of the technology to deliver it.

The assessment structure contains the following elements:

1. What the automation is intended to provide in terms of the level of control to be applied to the system
2. The capability of the technology in terms of:
 - a. the extent of the inputs: a simple system will only use one or two, a more complex system can theoretically take in any amount or type of input
 - b. the degree of freedom in the decision outcome: from the simplest “if these criteria are met, this action will be performed (only one outcome), through to multiple options of output, to any output
 - c. the complexity of the problem-solving element.
3. The operational context in which the automation will be implemented:
 - a. Passive: where the environment does not change in a way that requires the automation to adapt
 - b. Active: where the environment changes, typically as a result of the actions of other actors within the environment who have their own intent(s)

- c. Reactive: where the actions of other actors within the environment are actively attempting to affect the outcome of the automation (typically this would be a military environment)
- 4. A gross assessment of risk, taking into consideration the risk associated with:
 - a. the allocation of control
 - b. the type of task
 - c. the certainty of the information feeding the automation

4 Discussion of Problem Space

Taking a conceptual leap that this structure could be populated and could provide a useful framework, then the consequences and insights that this may bring can be considered. The following subsections look at each aspect in turn. They include the tables generated to form the assessment structure: these tables were generated during a project looking at the implementation of automated support to decision-making in a time-critical, safety-critical, complex environment. As such, the narrative descriptions may be more applicable to some domains than others: these tables are provided as examples of the assessment approach under consideration.

4.1 Statement of Intent

Looking at the intent of the automation in terms of what aspects are controlled by the system and what are controlled by the human, the following table presents a possible structure for generating a detailed description of the intent for the autonomous system. It is structured against an automation description loosely based on the Autonomy Levels developed by the Society of Automotive Engineers (SAE) [14], supplemented and adapted to make it applicable to autonomous systems across a variety of domains, not just automotive.

The result of completing this assessment table effectively forms a description of the allocation and/or supervision of tasks or types of tasks that the entire system (human and automation) intends to perform.

Table 5. Statement of Intent

| Level | Name | Narrative Definition |
|-------|---------------|--|
| 0 | No Automation | The human will perform all of the dynamic operational tasks. |
| 1 | Assistance | A single assistance system will utilise information about the environment to perform a mode-specific task: the human will perform all remaining aspects of the dynamic task. |

| Level | Name | Narrative Definition |
|-------|-----------------------------|--|
| 2 | Partial Automation | One or more assistance systems will utilise a range of information about the environment to perform mode-specific tasks: the human will perform all remaining aspects of the dynamic task. |
| 3a | Conditional Automation | A mode-specific automated system will perform aspects of the dynamic task with the expectation that the human will respond appropriately to a request to intervene |
| 3b | Conditional High Automation | A mode-specific automated system will perform all aspects of the dynamic task with the expectation that the human will respond appropriately to a request to intervene |
| 4 | High Automation | A mode-specific automated system will perform all aspects of the dynamic task, even if a human does not respond appropriately to a request to intervene |
| 5 | Full Automation | An automated system will perform all aspects of the dynamic task. |

4.2 Capability of Technology

Taking the statement of intent generated by the use of the first assessment table, the extent of the intended automation can then be considered. A modified OODA loop was considered to be a useful structure for this; presented in Table 6 below.

Using this framework to try to identify examples of technology that satisfy those technological requirements, it becomes evident (reasonably quickly) that for levels A to C there are existing examples to build upon: for example, a satnav could be categorised as Level B Advice and an automated system for reverse parking a car could be categorised as Level B Act. However, D and E have no tangible current examples outside of science fiction.

Table 6. Extent of the Intent of the Automation

| | Type | Definition | | |
|---|-----------------|--|--------|---|
| A | Observe | Monitors conditions and provides information in respect to them | Advice | Limited set of Inputs and single (or binary) output |
| | Observe & React | Monitors conditions and reacts when a condition (or Limited set of conditions) are met | Act | |
| B | Orient | Applies the monitored information to the situation | Advice | Limited set of Inputs and outputs |

| | Type | Definition | | |
|----------|------------------------------------|--|---------------|--|
| | Orient & Perform Action | Applies the monitored information in relation to the situation to perform a pre-specified activity when a condition (or set of conditions) are met | Act | direct logic - "Proveable" logic has "perfect/optimised" solution |
| C | Decide-Select | Applies the monitored information in relation to the situation to make a decision about the parameters of an activity: which activity, when or where. | Advice | Defined set of Inputs and outputs direct logic - can "not Proveable" logic |
| | Select/Act feedback loop | Applies the monitored information in relation to the situation, makes a decision about an activity (which activity, when or where) and performs that activity. | Act | Does not have a "perfect/optimised" solution. |
| D | Decide - Solve | Resolves multiple consequences and benefits | Advice | Defined set outputs but any inputs |
| | Solve/Act feedback loop | Applies the resolution of the multiple consequences and benefits by actioning - actioning the best options from it output set. | Act | Resolves consequences and benefits and selects actions with acceptable levels of risk across multiple dimensions |
| E | Decide - Solve | Resolves multiple consequences and benefits | Advice | Any Inputs and Any outputs |
| | Solve/Act feedback loop | Applies the resolution of the multiple consequences and benefits by actioning: <ul style="list-style-type: none"> - the positive steps - the mitigations that reduce the likelihood of negative effects - the mitigations that increases the likelihood of the process staying in the optimum range | Act | Resolves consequences and benefits and selects actions with acceptable levels of risk across multiple dimensions |

4.3 Risk Assessment

Risks to the successful implementation of autonomous systems can be associated with:

- the allocation of control and the interaction between the human and the automation
- the type of task and the operational context
- the uncertainty of information being used to feed the automation.

A set of risk assessment profiles have been constructed to represent each of these types of risk for each logical combination of intent and technology capability: there are nineteen of these logical combinations, leading to a total of 57 risk assessment profiles. The data within these profiles has been derived from currently available research in this domain together with human factors best practice. They are structured to allow new research findings to be included as and when identified. Some of the profiles are currently only sparsely populated.

The objective of this risk assessment process is to provide useful guidance for the implementation of an appropriate level of automation. The profiles allow a quick identification of when a proposed intent combined with a proposed technology level would be “high risk”: a simple example would be that a high level of uncertainty of input data would indicate a high risk if used for a highly automated system and therefore would imply the need for a human presence within the implementation of the system.

The aspects of risk covered by these profiles are reasonably well understood and are founded in current research. However, in the examples assessed it has been determined that there is a tendency that as the domain becomes more uncertain, the complexity of the solution increases, which increases the risk of insufficient time for a human to interact with the system. This leads to a new category of risk, which is related to the relation between the increasing complexity of automation and its interface to humans: defined as the “logos chasm”.

4.4 The Replacement of Human Cognition with Automation

As mentioned above, technology has (to date) failed to produce any examples of higher level autonomy. These higher levels tend to have a component of decision making, which requires a cognitive process.

One approach to examining this problem is to compare the detailed capability of technology to the cognitive processes it is trying to replace. Bloom’s taxonomy [15], although conceived as a guide and tool for education, provides a useful structure for comparison.

Most representations of Bloom's taxonomy present it as a two-dimensional matrix with the dimensions of cognitive process and knowledge. The cognitive process dimension consists of:

- Create - Put elements together to form a coherent whole; reorganise into a new pattern or structure
- Evaluate - Make judgments based on criteria and standards
- Analyse - Break material into constituent parts and determine how parts relate to one another and to an overall structure or purpose
- Apply - Carry out or use a procedure in a given situation
- Understand - Construct meaning from instructional messages, including oral, written and graphic communication
- Remember - Retrieve relevant knowledge from long-term memory

and the knowledge dimension consists of:

5. Metacognitive - Knowledge of cognition in general as well as awareness and knowledge of one's own cognition
6. Procedural - How to do something, methods of enquiry, and criteria for using skills, algorithms, techniques and methods
7. Conceptual - The interrelationships among the basic elements within a larger structure that enable them to function together
8. Factual - The basic elements that must be known to be acquainted with a discipline or solve problems within it

When considering computational technology the limitations of a Turing machine [16] must be understood. A Turing machine can be defined as "a mathematical model of a hypothetical computing machine which can use a predefined set of rules to determine a result from a set of input variables".

Although computer scientists are looking at a variety of hyper-computing solutions: i.e. computing solutions that extend past the confines of a Turing machine; it is true to say that currently, all computers have a Turing machine restriction. Until hyper-computing becomes available, automation can only represent a logical rule-based system. In attempting to create more completely autonomous systems, this means that the logic of the system becomes increasingly complex to try to compensate for the lack of both metacognition and creative imagination.

Without going in to the detailed mathematical proofs of this, it can be shown that a Turing machine cannot solve the create level of cognitive processing or access the metacognitive level of knowledge. To put it a different way, it is possible to envisage a logical set of rules and process flows that satisfy the bottom five cognitive and bottom four knowledge dimensions of Bloom's taxonomy. This apparent separation between the Turing logic (logos) and the creative human processes (imagination) has reverberations of Jung's concept of "logos" and "imagination" [17]: which he then equated to conscious thought and subconscious thought. Taking this analogy to its

extreme, current computing capabilities only utilise conscious thought, it is well-established that during the decision-making process, humans frequently utilise both conscious and subconscious thought.

This realisation goes a considerable way towards explaining why there are no examples of current technology at levels D and E as noted above. It may also imply that until hyper-computing becomes available, that boundary cannot be crossed (the logos chasm).¹

4.5 Consequences for Autonomous Systems

When considering autonomous systems, there are three dimensions to explore:

- The quality of interaction over the task split between human and machine: this leads to discussions on the quality of a human decision versus a machine decision, which in turn implies the need for an understanding of the difference between machine logos and human logos and imagination. The question is, is there a notion of machine intelligence (which is all logos) which could be equally as good as (but different to) human intelligence?
- The time dependency of task split between human and machine: given that real life tasks happen in a time-constrained situation, and that machines are capable of computational calculations from sensors faster than humans can process the data, at what point does the fallback that humans can override the system become invalid? This identifies the point at which the human does not have the capability to react within the time constraints of the task and be able to undertake a comparably good “calculation” and generate a decision.
- The potential benefits of task split between human and machine: given that some cognitive tasks span areas of creative thinking and metacognition, how can the task be broken apart in order to allow the human to contribute in those ways, while gaining the benefit of the machine doing the computational calculations faster than the human can?

An additional cross-cutting aspect is that many real life situations are “open loop”: i.e. once a decision has been made, it is impossible to tell if it was in fact the optimum decision as, unlike in a laboratory experiment, the outcomes of different decisions can be compared.

It seems sensible to utilise the unique abilities of both the human and the automation to their most beneficial extent within the entirety of the system. However, for automated systems that are trying to replicate decision processes in complex environments, this raises some critical questions, such as:

1. for a system with a large number of sensors and a large number of logical outcomes (decisions that could be made, given the information provided by the sensors) and a complex set of rules, how can it be organised so that an

¹ Extensive fuzzy logic engines can solve the “end problem”, but only by virtue of the fuzzy logic’s definition assuming that the fuzzy logic will always resolve.

individual can judge the decision made (or recommended) by the automation?

2. accepting that computers are both more accurate and faster than humans at processing data (given the appropriate ruleset) how can it be arranged for an individual to be able to monitor the logical process followed by the computer to the extent that they can identify cases where they should override it?

The logical consequences of these questions are at one extreme, the human always has to accept the autonomous system's decision and never overrule it, and at the other extreme, the autonomous system has to be limited by the ability of the human to judge its decisions.

This logical barrier (referred to as the logos chasm) can be considered in three different ways:

- in terms of a barrier to intent
- in terms of a barrier to implementation
- in terms of a barrier to replacing human cognitive processes.

The logos chasm is effectively the gap between what is currently achievable in terms of the implementation of autonomous systems and what requires hyper-computing for its solution.

This does not mean that automation that sits inside this logical chasm should not be put into place; what this construct means is that when it is recognised that automation is sitting within that chasm space, mechanisms must be put into place to mitigate for the consequences.

5 Summary and Conclusions

Examining the multitude of papers that consider automation frameworks from different points of view, they relate autonomy to advice that roughly falls into the categories of:

- the allocation of control and the interaction between the human and the automation
- the type of task
- the uncertainty of information being used.

However no one framework seems to explicitly cover them all. They intuitively include the issues but do not always state why these are issues and when they are going to trigger, or relate them to the capabilities of the technology.

Additionally, they do not explicitly distinguish when autonomy steps into this chasm space, and therefore do not explicitly relate any mitigations or advice to the

understanding that this chasm is currently unbridgeable and has the known ramifications identified above.

The framework discussed in this paper is intended to compensate for some of the inadequacies of individual approaches to the implementation of autonomous systems by providing an overall structure that includes both an assessment of the associated risks and an assessment of the barriers to successful implementation of the system.

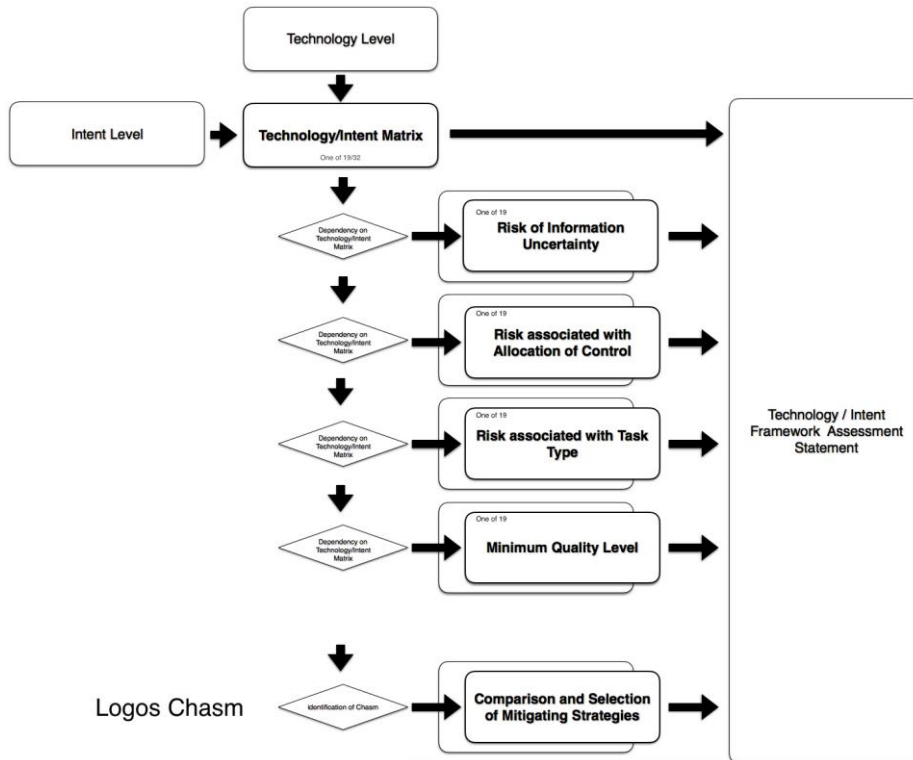


Fig. 1. Illustration of Overall Structure

Although there are currently some concepts related to mitigating the logos chasm, these concepts have not been extensively investigated. Therefore future work is planned to attempt to identify and validate strategies that will mitigate the consequences of this chasm space.

It is hoped that this identification of the logos chasm and the multi-faceted risk assessment approach presented in this paper will provide a useful construct for implementing acceptable autonomous systems.

References

1. Endsley, M.: The application of human factors to the development of expert systems for advanced cockpits. In: 31st Annual Meeting of Human Factors Society, pp. 1388e1392. (1987)
2. Sheridan, T. B.: Telerobotics, Automation, and Human Supervisory Control. Cambridge, MA: MIT Press. (1992)
3. Parasuraman, R.: Designing Automation for Human Use: Empirical Studies and Quantitative Models. *Ergonomics* 43, no. 7: 931–951 (2000).
4. Proud, R. W., Hart, J. J.: FLOAAT, A Tool for Determining Levels of Autonomy and Automation, Applied to Human-Rated Space Systems. Infotech@Aerospace, American Institute of Aeronautics and Astronautics (AIAA) 2005-7061. Arlington, VA: AIAA. (2005)
5. Osinga, F.: Science, Strategy and War: The Strategic Theory of John Boyd. Eburon Academic Publishers, Delft, The Netherlands (2005)
6. Bannon, L. Reimagining HCI: Toward a more human-centered perspective. *interactions*, 18(4), 50-57 (2011).
7. Kun, A. L., Boll, S. and Schmidt, A.: Shifting Gears: User Interfaces in the Age of Autonomous Driving. *IEEE Pervasive Computing*, 15 (1), 32-38, Jan.-Mar. 2016. doi: 10.1109/MPRV.2016.14 (2016)
8. Cramer, H. S., Evers, V., Van Someren, M. W., & Wielinga, B. J.: Awareness, training and trust in interaction with adaptive spam filters. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 909-912). ACM. (2009)
9. McCarthy, J.: Bridging the gaps between HCI and social media. *Interactions*, 18(2), 15-18(2011).
10. Cresci, E.: Chatbot that overturned 160,000 parking fines now helping refugees claim asylum. *The Guardian*. <https://www.theguardian.com/technology/2017/mar/06/chatbot-donotpay-refugees-claim-asylum-legal-aid> (2017)
11. Saunder, J., Hunt, P. & Hollywood, J.S.: Predictions put into place: a quasi-experimental evaluation of chicago's predictive policing pilot. *Journal of Experimental Criminology*, 12 (3) 347-371 (2016)
12. Caringi, A.: The FDA Approved The World's First Autonomous Insulin Pump Vocativ: Health <http://www.vocativ.com/363955/the-fda-approved-the-worlds-first-autonomous-insulin-pump/index.html> (2016)
13. Platania, R., Shams, S., Yang, S., Zhang, J., Lee, K. and Park, S-J.: Automated Breast Cancer Diagnosis Using Deep Learning and Region of Interest Detection (BC-DROID). In Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB '17). ACM, New York, NY, USA, 536-543. DOI: <https://doi.org/10.1145/3107411.3107484> (2017)
14. J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. SAE International: (2014)
15. Bloom, B. S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H.; Krathwohl, D. R.: Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. David McKay Company, New York (1956)
16. Shieber, S. M.: Lessons from a restricted Turing Test. *Communications of the ACM*, 37(6):70–78. (1994)
17. Shelburne, W. A.: Mythos and Logos in the Thought of Carl Jung: The theory of the Collective Unconscious in Scientific Perspective. State University of New York Press, Albany N. Y. (1988)